# Cost-Effective Accurate 3-D Reconstruction Based on Multi-View Images for Plant Phenotyping

**Lu Lou**

Department of Computer Science
Aberystwyth University

June 2016

This thesis is submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

# Declaration

This thesis has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed .....(candidate)

Date .....

# Statement 1

This thesis is the result of my own investigations, except where otherwise stated.
Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed .....(candidate)

Date .....

# Statement 2

I hereby give consent for my thesis, if accepted, to be made available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed .....(candidate)

Date .....

# ABSTRACT

Phenotyping involves the measurement, ideally objectively, of characteristics or traits, usually in the context of living organisms, including plants. Traditionally, this is limited to either tedious and sparse manual measurements, often acquired destructively, or coarse image-based 2-D measurements. A dynamic model of 3-D architecture through developmental time could capture useful geometric characteristics representing phenotypic information on morphology and also record the adaptive response to environmental conditions. In recent years, many emerging 3-D imaging technologies (based on laser scanners, structured light, multi-view stereo, etc.) provide the potential to capture quantitatively morphological features and have been proposed as non-destructive phenotyping alternatives to current cost-intensive commercial phenotyping platforms. However, available 3-D resolutions are limited in various ways. For example, they may focus on a specific organ (e.g. leaf or stem) or tend to be qualitative rather than providing quantitative information and often lack any estimation of accuracy. This thesis investigates the existing methods and aim to build a cost-effective accurate 3-D reconstruction framework that could cope with a diversity of plant forms and sizes, while using equipment available to most biology labs. I firstly developed a multi-view image acquisition method to obtain high-resolution 2-D image sequences, using low-cost turntable and a consumer-grade digital camera. To deal with the self-occlusion problem, the proposed method is based on short baseline and multiple viewpoint photography, and usually acquired approximate 60-120 images. To improve the flexibility and save time of data acquisition, the camera is equipped with a variable focal length lens and does not need extra calibration processing. I proposed an efficient Structure-From-Motion method to estimate cameras' parameters and poses from the uncalibrated images of the plant. Finally, I developed an accurate multi-view stereo 3-D reconstruction method to yield a dense and detailed 3-D point cloud of plant. The method takes both accuracy and efficiency into account, and therefore is faster than other state-of-the-art methods. The accuracy of the proposed method was evaluated with limited ground truths as well as the mea-

surement deviations of test specimens. The experimental results show that the proposed methods are highly reliable for phenotyping of various plants during their entire growth cycles. This promising generalized 3-D imaging technique has the potential to be implemented in automated procedure.

# ACKNOWLEDGEMENTS

I would like to express my uttermost gratitude to my supervisors: Dr. Yonghuai Liu and Prof. John Doonan, for their motivation, enthusiasm, and guidance, which have been essential at all stages of my research.

I would also like to thank the members of the National Plant Phenomics Centre of IBERS of Aberystwyth University, especially Dr. Fiona Corke, Mr. Alan Gay, Dr. Jiwan Han, Dr. Candida Nibau, Dr. Kevin Williams, Andreu Alcalde-Barrios, Robert Bellow, for kind supports on acquisition of plant materials.

I want to thank Dr. Frèdèric Labrosse, Dr. Mark Neal and Dr. Hannah Dee, for their guidance and help on my early studying. I am also very grateful to the members of Vision, Graphics and Visualisation group, especially Prof. Reyer Zwiggelaar, Dr. Bernie Tiddeman, for their feedback, questions, and suggestions on my research. Many thanks also go to the members of Intelligent Robotics group, especially Dr. Colin Sauze, Michael Clarke, Peter Scully, Marek Ososinski, Suzana Barreto, Tom Blanchard, and the other staff in the Department of Computer Science, for their friendliness and innumerable help.

My special thanks go to Prof. Qiang Shen, it is impossible to start my PhD study in Aberystwyth University without his introduction and recommendation. I am very grateful to my home university, Chongqing Jiaotong University, that provide me with the support of studying.

I would also like to thank all my friends, especially the Chinese student community in Aberystwyth for their continuous support.

Finally, my heart overflowed with gratitude to my family: my father Jun Lou, my brother Zhi Lou, Qian Lou and Yan Lou, my sister Hua Lou, my dear wife Xin Xu and my dear daughter Xinyu Lou. The completion of this thesis would not have been possible without their constant support and encouragement.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

This chapter provides a brief introduction on the significance of plant breeding and the challenge of plant phenotyping, then the motivation and objective of the study are introduced, and this chapter concludes with an outline of the thesis.

## 1.1 Plant Breeding and Phenotyping

Agriculture is facing tremendous challenges to feed the world population. The global population is expected to increase from ∼6 billion in 2000 to 9.1 billion in 2050 (Tester & Langridge, 2010). Global production of three major cereal crops, rice (Oryza sativa), maize (Zea mays) and wheat (Triticum aestivum), must be increased by at least 70% before 2050 to meet the demand of increasing global population, whereas the food supply has already become insufficient (Furbank & Tester, 2011). In addition, the demand for biofuels will increase (Sticklen, 2007), and the competition for the limited arable land between food and biofuel crop plants makes food security challenges even worse. Unfortunately, as a consequence of global climate change, crop production is already more frequently exposed to extreme weather, faces limited availability of water and nutrient resources, and will be restricted.

To address these issues, we need new high-yielding genotypes of agricultural crops adapted to the future environment and climate.

### 1.1.1 Plant Breeding

Plants, as renewable natural resources, are the main source of human food, medicine, bio-energy, animal feed and other raw materials. Most of crops that dominate the agricultural landscape, were developed during the agricultural revolution when human societies made the transition from nomadic hunter-gatherers to horticulture and agriculture. These crops largely support human nutrition. For example, cereal and rice, which are rich in carbohydrates, provide the staple of many human diets; beans and nuts supply proteins; fruits and vegetables provide nutrients, vitamins, and fibre; fats are derived from crushed seeds, peanut and rapeseed oils, or fruits.

Crop improvement has been practiced for thousands of years, since the beginning of human agriculture, eight to ten thousand years ago (Sleper & Poehlman, 2006). Early farmers noticed that not all plants were the same – some plants may have grown larger than others, or maybe some grains tasted better or were easier to grind. The farmers saved those seeds from plants with desirable characteristics and planted them for the next season's harvest. This process is known as artificial selection. Once the science of genetics became better understood in the early 20th century, plant breeders used what they knew about genetics to select for specific desirable traits to develop improved varieties. The selection for features such as faster growth, higher yields, pest and disease resistance, larger seeds, or sweeter fruits has dramatically changed domesticated plant species compared to their wild relatives (ISAAA, 2006). Plant breeding can be considered as the purposeful manipulation of changing the traits of plants in order to produce desired characteristics.

### 1.1.2 Plant Phenotyping

The terms *phenotype* and *genotype* were introduced by the Danish plant scientist Wilhelm Johannsen in 1909 in his textbook titled *The Elements of an Exact Theory of Heredity* (Johannsen, 1903), and were developed more fully in his paper titled *The Genotype Conception of Heredity* in 1911 (Johannsen, 1911). The concepts of genotype and phenotype were an outgrowth of Johannsen's pure-line breeding experiments on barley and the common bean. Johannsen's

genotype-phenotype distinction explicitly prescribed a relationship between the study of development and the study of heredity. According to Johannsen, the genotype of an organism gives rise to the organism's phenotype through the process of development, under the influence of the environment.

Phenotype can be broadly defined as an organism's expressed physical traits. Examples of an organism's phenotype include traits such as color, height, size, shape, and behavior, etc. Fig. 1.1 shows some phenotypes of the legumes, include pod color, pod shape, pod size, seed color, seed shape, and seed size. The phenotypes of all organisms are determined by their genotype, the environment, and their interaction (Johnson & Agrawal, 2005).

The genotype refers to the gene (and allele) composition of an organism or cell for a given trait. The genotype of an individual is determined by its parents. In sexually reproducing species, the genotype is a blend of the 2 parents and this recombination can produce diversity in the offspring. It is this diversity that breeders need to measure and then select favorable variants.

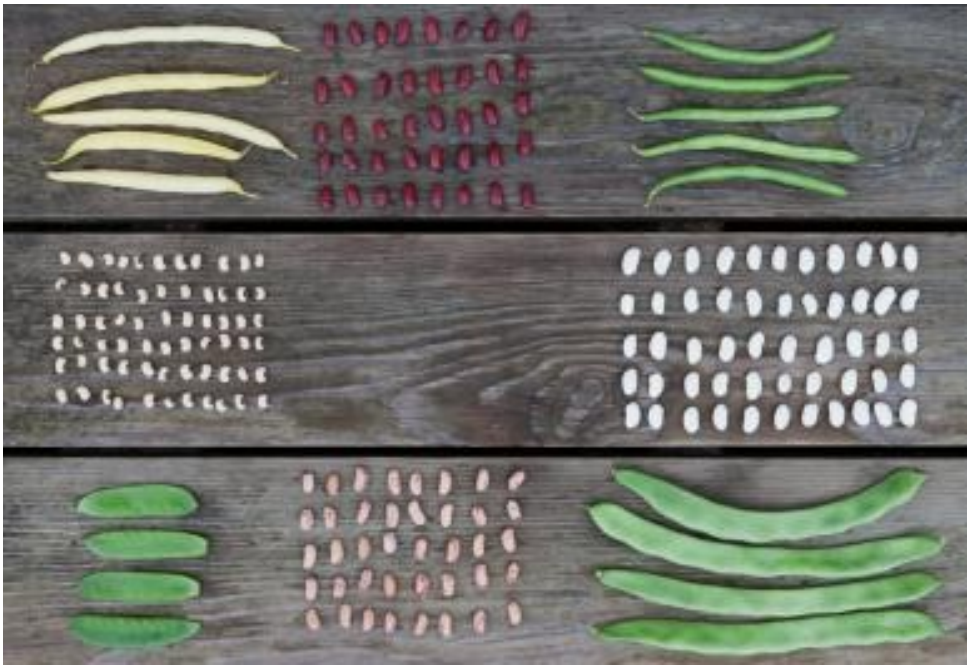


Figure 1.1: Phenotypes expressed in legumes image include pod and seed color, size, and shape. Image courtesy of (Bailey, 2014)

In plant breeding and quantitative genetics, usually hundreds or even thousands of measurements are performed to select superior individuals or identify

regions in the genome controlling a trait (Walter *et al.*, 2015).

The "phenome" can be considered as the set of phenotypes expressed by a cell, tissue, organ, organism, or species, and represents the sum of its phenotypic traits. Phenomics is an area of biology concerned with the measurement of phenomes the physical and biochemical traits of organisms as they change in response to genetic mutation and environmental influences (Schilling *et al.*, 1999).

The increasing capabilities of analytical chemistry has broadened the concept of a quantitative analysis of traits to the description of the variability of proteins (Schulze & Usadel, 2010), of metabolic pathways (Schauer & Fernie, 2006) and of other characteristicss of living plants (Walter *et al.*, 2015).

Plant *phenotyping*, therefore, is the comprehensive assessment of plant complex traits such as growth, development, tolerance, resistance, architecture, physiology, ecology, yield, and the basic measurement of individual quantitative parameters. Examples for such direct measurement parameters are image-based projected leaf area, chlorophyll fluorescence, stem diameter, plant height/width, compactness, stress pigment concentration, tip burn, internode length, colour, leaf angle, leaf rolling, leaf elongation, seed number, seed size, tiller number, flowering time, germination time, etc.(Walter *et al.*, 2015)(LemnaTec., 2014)(Granier *et al.*, 2006)(Reuzeau *et al.*, 2006).

It is clear that the phenotype is characterized by an enormous amount of processes, functions and structures, which are changing during growth and development (Walter *et al.*, 2015). Fig. 1.2 illustrates a genotype-environment interaction (or Genotype-by-Environment, $G \times E$), where two different genotypes respond to environmental variation (such as limited resources of environment A vs. B.) in different ways, this genotype-by-environment interaction results in different phenotypes which are observable at various organizational levels. In this sense, phenotyping can be considered as far more complex than the analysis of the linear arrangement of genes in the genotype (Houle *et al.*, 2010) and, in practice, it is impossible to fully characterize the phenome.

Understanding the linkage between a particular genotype and a specific phenotypic parameter is a core goal of modern biology. However, it is generally difficult due to the large number of genes and the interaction with complex and

changing environmental influences (Houle *et al.*, 2010).

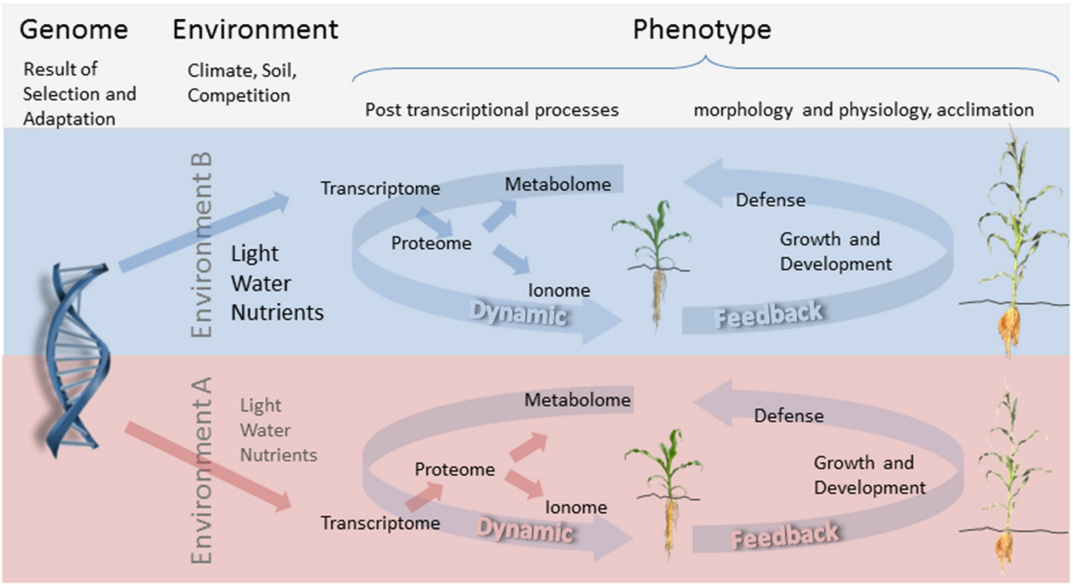



Figure 1.2: Relation between genotype and phenotype. Image courtesy of (Walter *et al.*, 2015).

For a typical plant phenotyping experiment, plants are grown in controlled conditions (growth chambers or glasshouses) and subjected to different environmental conditions and stresses (e.g. drought, salt, temperature, etc.) with the primary aim of monitoring their phenotypic response using various measurements (Granier & Tardieu, 2009)(Schurr *et al.*, 2000). Common morphological traits of interest include parameters such as main stem height, size and inclination, petiole length and initiation angle, and leaf width, length, inclination, thickness, area, and biomass (Furbank & Tester, 2011)(Granier & Tardieu, 2009)(Schurr *et al.*, 2000)(Vos *et al.*, 2010). The traditional procedure to collect these data consists of many laborious manual measurements (Fig. 1.3, Fig. 1.4), often requiring destructive harvest and thus multiple replicates of individual plant genotypes or varieties to allow successive harvests over time. A typical manual phenotypic analysis of 200 plants require 100s man-hours of work (Paproki *et al.*, 2012).

Figure 1.3: Manual measurement of a maize in the field.

Figure 1.4: Manual measurement of a maize in the glasshouse.

### 1.1.3 *Arabidopsis thaliana*

The vascular plant *Arabidopsis thaliana* is a small flowering plant native to Eurasia. *Arabidopsis thaliana* has many advantages, including a short generation time, small size, large number of offspring, and a relatively small nuclear

genome. These advantages promoted the growth of a scientific community that has investigated the biological processes of *Arabidopsis* and produced a high quality genome sequence (Initiative, 2000).

*Arabidopsis* is a popular model organism in plant biology and genetics. Fig. 1.5 illustrates the life cycle of *Arabidopsis* (Krämer, 2015), where Fig. 1.5(A) illustrates the accession Columbia (*Col*) at different stages of its life cycle, from seed (bottom left) to seedling (11 days), to vegetative growth (39 days), and to reproductive growth (45 days); Fig. 1.5(B) shows photographs of a flower; Fig. 1.5(C) is a pollen grain (scanning electron micrograph), and Fig. 1.5(D) presents mature siliques (seed pods; left: closed; right: open with a few remaining unshattered seeds) at higher magnification.

### 1.1.4 Modern Plant Phenotyping System

Phenotyping is the major bottleneck limiting the progress of genetic analysis and genomics prediction (Tardieu, 2009).

For modern phenotyping, an important requirement is to improve the accuracy (how close the process or measurement is to the absolute truth), precision (the repeatability or variance of the measurement process) and throughput of phenotyping estimation at all levels of biological organization, while reducing costs and minimizing human labour by means of automation, integrated techniques and experimental design (Cobb *et al.*, 2013).

In recent years, various automated high-throughput phenotyping platforms have been developed, PHENOPSIS (Granier *et al.*, 2006) is used by French National Institute for Agricultural Research (INRA) for Arabidopsis. TraitMill™ (Reuzeau *et al.*, 2006) is developed by the company CropDesign used on evaluation of transgenic rice (Oryza sativa). Commercial high-throughput phenotyping platforms (LemnaTec., 2014) based on automated plant handling and imaging systems have been installed at Australian Centre for Plant Genomics (ACPFG), the Leibniz Institute of Genetics and Crop Plant Research (IPK) in Gatersleben, and The National Plant Phenomics Centre (NPPC) of the Institute of Biological, Environmental and Rural Sciences (IBERS) of Aberystwyth University in the UK. In such systems, morphological and physiological data are captured by a variety of sensors: visual infrared, near infrared and fluorescence

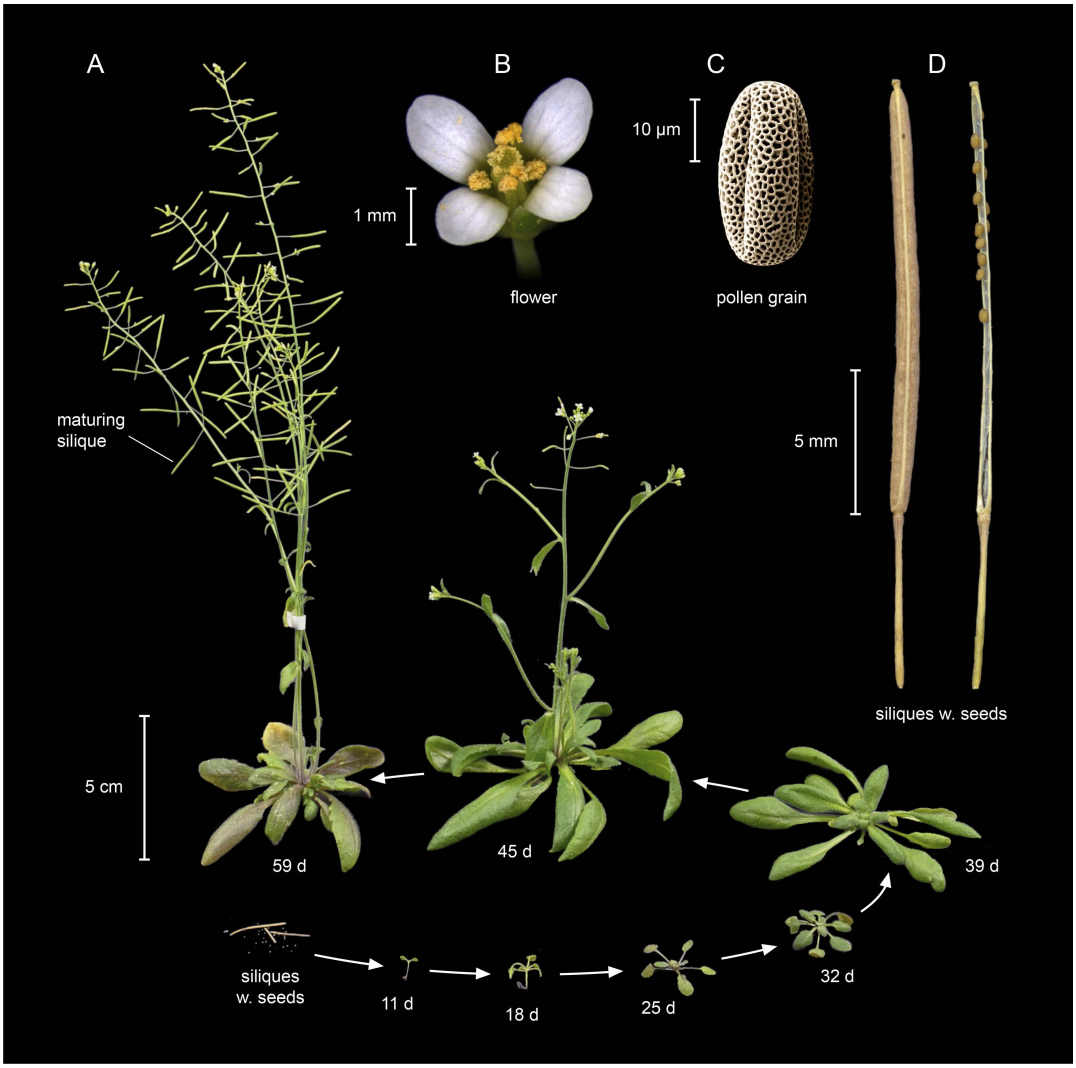



Figure 1.5: Life cycle of *Arabidopsis thaliana*. (A) illustrates the accession Columbia (*Col*) at different stages of its life cycle, from seed (bottom left) to seedling (11 days), to vegetative growth (39 days), and to reproductive growth (45 days); (B) shows photographs of a flower; (C) is a pollen grain (scanning electron micrograph), and (D) presents mature siliques (seed pods; left: closed; right: open with a few remaining unshattered seeds) at higher magnification. Image courtesy of (Krämer, 2015).

(LemnaTec., 2014).

These high-throughput phenotyping platforms are aiming at reducing manual acquisition of phenotypic data but do so at considerable high cost.

#### 1.1.4.1 PHENOPSIS

The PHENOPSIS platform (France, 2014) consists of 3 automatons (PHENOPSIS 1, 2 and 3) that function to weigh, accurately irrigate and digitally image (visible and infrared) individual *Arabidopsis thaliana* plants growing under rigorously controlled environmental conditions. The automatons are located in 3 individual growth chambers. PHENOPSIS is associated with a database, PhenopsisDB, which organizes all measurements, images and metadata. In each experiment, one pot is covered with graph paper to calibrate the images. Images over the visible are analysed with *ImageJ* image analysis software (ImageJ, 2014) in order to get rosette expansion area over time. Infrared images allow the measurement of rosette surface temperatures.

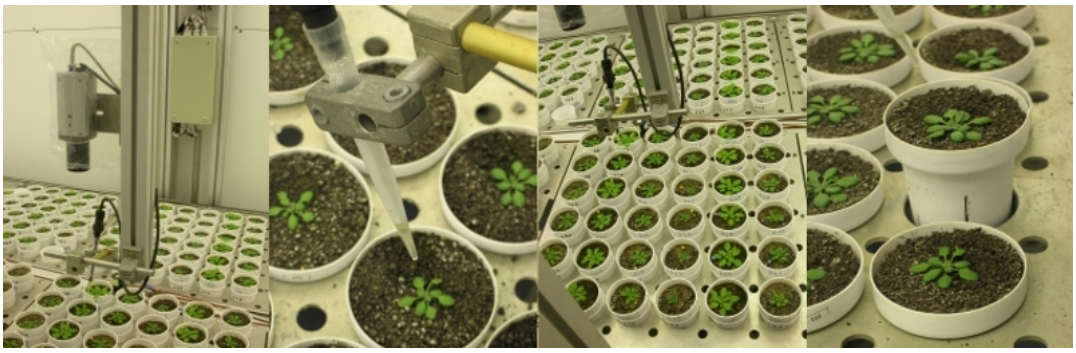


Figure 1.6: The PHENOPSIS platform. Image courtesy of (France, 2014)

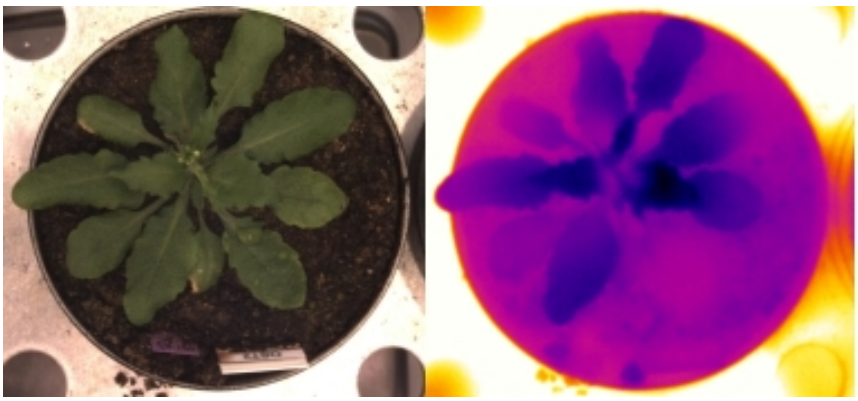


Figure 1.7: Visible and infrared images of Arabidopsis rosette in PHENOPSIS platform. Image courtesy of (France, 2014)

Main technical characteristics of PHENOPSIS platform are as follows:

(1) Installation for: small plants, *Arabidopsis*.

(2) Environmental monitoring: temperature, humidity, light;

(3) Parameters: leaf area, individual organs, transpiration (FIR), 3-D architecture

(4) Capacity: 1510 plants;

(5) Average experiment duration: 60 days.

#### 1.1.4.2 TraitMill™

CropDesign has developed TraitMill™ (Reuzeau *et al.*, 2006), an automated plant evaluation platform allowing high-throughput testing of the effect of plant-based transgenes on agronomically valuable traits in crop plants. The focus of the platform is currently on rice, a good model for other important cereals such as maize and wheat. TraitMill™ offers a high-throughput prediction of gene function. Genes of validated function that confer trait improvement can then be transferred to other cereal crop species such as maize, but also to dicots, trees and ornamentals.

#### 1.1.4.3 WPS's High-Throughput Plant Phenotyping

In 2003, WPS ("We Prove Solutions" Company) (WPS, 2014) constructed the world's first high-throughput phenotyping facility. Worldwide with 100+ installations across several continents, the automated high-throughput plant phenotyping platforms provided by WPS handle millions of plants every day.

Fig. 1.8 illustrates WPS's automated high-throughput plant phenotyping system, which consists of 4 parts: (1) The section of the conveyor is counterbalanced and simply lifts up, manually, for easy walk-through access to the imaging area and watering module; (2) Plant growth and conveyance area; (3) Imaging station, and (4) Watering and weigh station.

Fig. 1.9 presents its architecture of experiment control software.

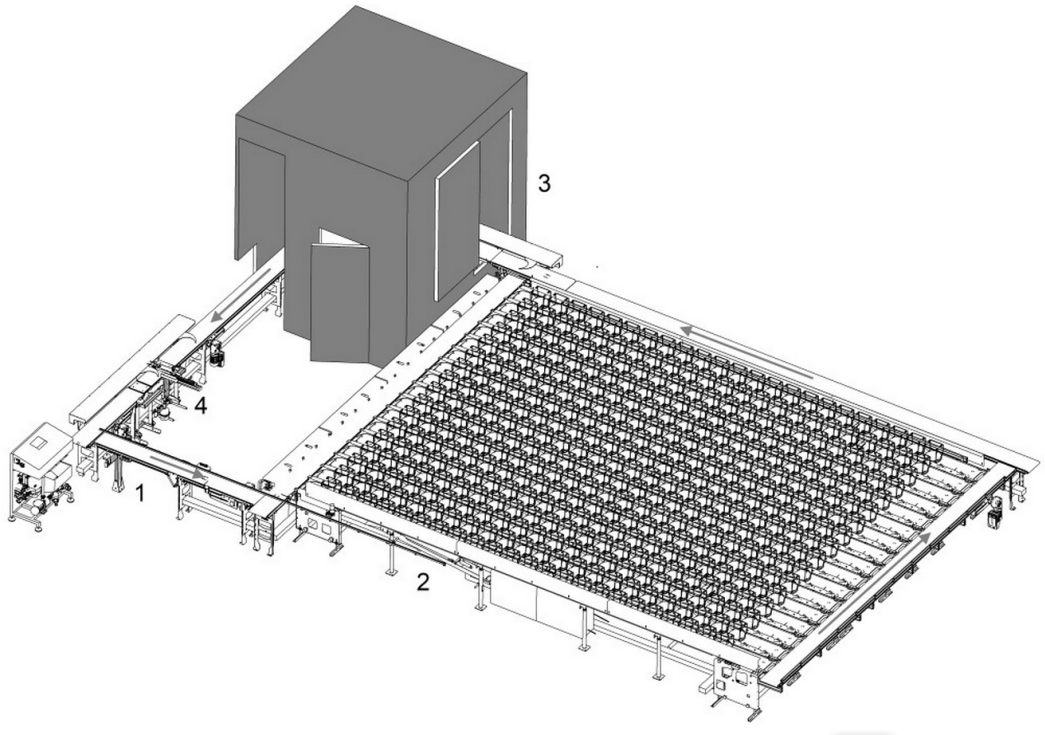



Figure 1.8: Plant automated conveyor system with multi-purpose imaging cabinets. Image courtesy of (WPS, 2014)

#### 1.1.4.4 LemnaTec's Plant Phenotyping Platform

LemnaTec (LemnaTec., 2014) offers a wide range of supported cameras and sensors to address divers plant phenotyping requirement. The architecture of platform of Lemntec is very similar to WPS's, but has 3 imaging cabinets with visiable RGB camera, Near infrared and ultra violet light imaging.

Each plant is located in a carrier which is situated upon a conveyor belt. The conveyor belt system automatically retrieves each plant as needed and passes it through the image capture units (see Fig. 1.10). In this system, plants are captured autonomously in near infrared-, ultra violet- and visible images in three boxes. Each box contains a top view camera and a side view camera and furthermore a turnable lifter which enables the plants to be lifted and turned. Fig. 1.11 presents its ability of multi-view imaging plants.

After imaging, plants pass the watering and weighing unit, which automatically measures weights and waters the plants. Since each carrier is tagged with an RFID chip each plant can be identified and traced during its growth cycle.

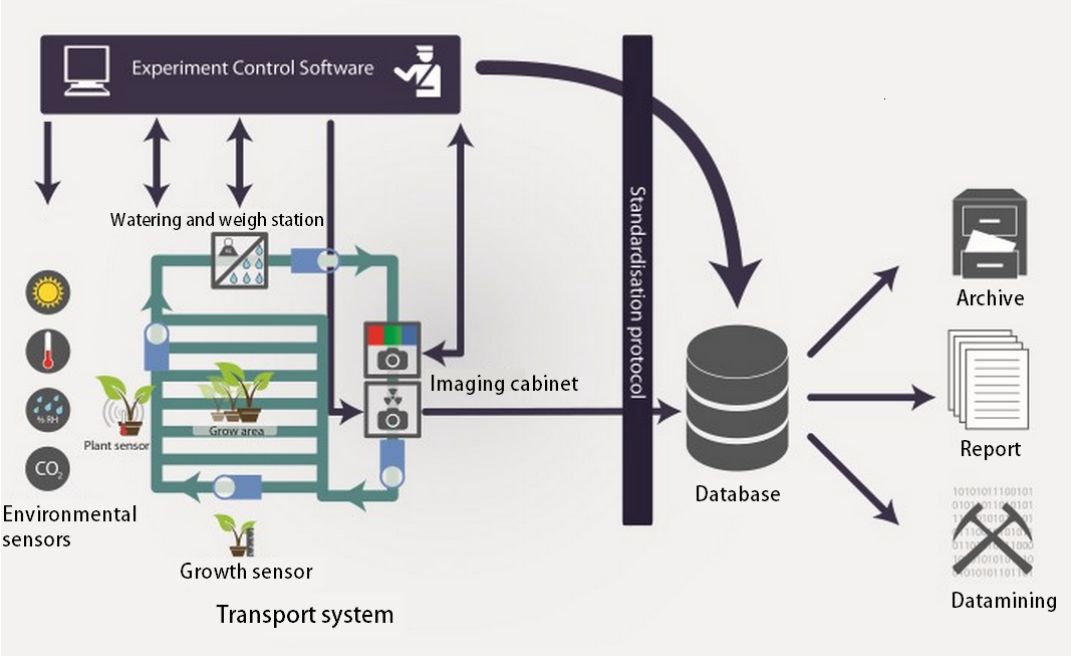



Figure 1.9: Architecture of experiment control software. Image courtesy of (WPS, 2014)

An extensive amount of data is generated by the platform within a period of plant development, and stored in a database system(Hartmann *et al.*, 2011).

LemnaTec platform is considered as the state of art high-throughput plant phenotyping.

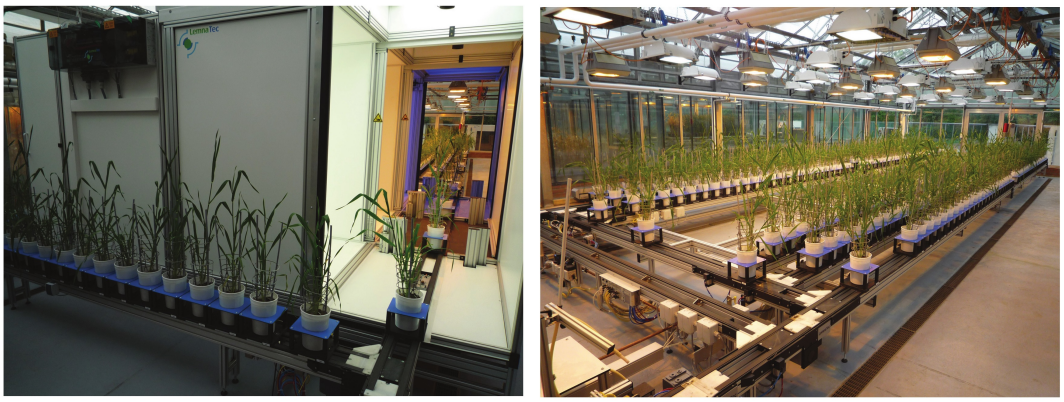


Figure 1.10: View of the greenhouse in LemnaTec platform. Image courtesy of (Hartmann *et al.*, 2011)

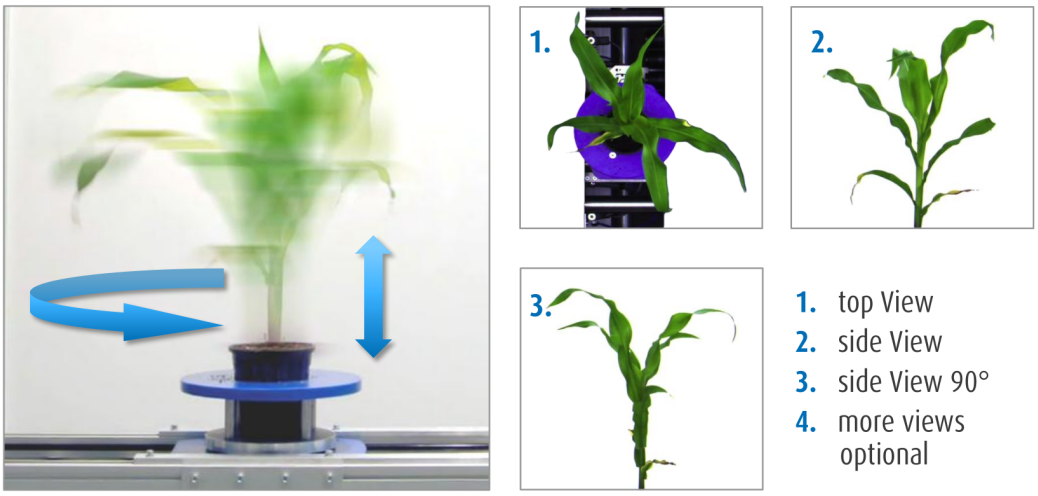



Figure 1.11: Multi-view imaging plants in LemnaTec platform. Image courtesy of (LemnaTec., 2014)

## 1.2 Problem Statement and Motivation

In recent years, digital camera, as a 2-D sensing equipment with advantages of low-cost and non-destructive, has been widely applied in plant visualization/-analysis techniques and high-throughput plant phenotyping platforms. 3-D imaging technologies (such as 3-D laser scanning, structured light and digital photography, etc.) are increasingly incorporated into mass produced consumer goods and have also the potential to be developed a resolution of plant 3-D acquisition.

### 1.2.1 Shortcomings of 2-D Imaging Techniques

Plant 2-D imaging (or digitization) has become an active research area, that aims to capture and measure the plant (or its organs) from digital images, and then carry out analysis. 2-D imaging and analysis techniques have been applied to tasks such as crop (or weed) detection and segmentation (Romeo *et al.*, 2013; Montalvo *et al.*, 2013; Meyer *et al.*, 2004; Onyango & Marchant, 2001; Zheng *et al.*, 2009; Onyango & Marchant, 2003; Samal *et al.*, 2006; Tellaeche *et al.*, 2008; Guillot *et al.*, 2009; Swain *et al.*, 2011), leaf and fruit recognition (Bruno *et al.*, 2008; Backes *et al.*, 2009; Nandi *et al.*, 2014; Zeng *et al.*, 2009), and plant

disease and pest analysis (Xu *et al.*, 2011). Fernandez *et al.*(Fernández *et al.*, 2013) presented an automatic system that combines RGB and 2-D multispectral imagery for discrimination of Cabernet Sauvignon grapevine elements such as leaves, branches, stems, and fruits in natural environments, to serve precision viticulture. Bac *et al.*(Bac *et al.*, 2013) used multi-spectral imaging to capture features of plants in a greenhouse, and then separate vegetation images into different parts for the aim of creating a reliable obstacle map for robotic harvesting. Li *et al.*(Li *et al.*, 2014a) developed an efficient method to identify blueberry fruit of four different growth stages in color images acquired under natural outdoor environments, and they quantitatively evaluated the results across several classifiers for fruit identification.

2-D imaging techniques are easy to implement because of their low-cost and availability. Nevertheless, some inherent shortcomings deeply restrict their applicability:

(1) The imaging and measuring process is subject to environmental illumination, such as controlled lighting or complex background;

(2) Pre-calibration (or marker setup) are always required;

(3) 2-D image does not carry the depth (range) information about the plant, hence is unable to recover the spatial distribution.

These shortcomings can be explained by a real study case, Leaf Segmentation and Counting Challenges, which organised in the workshop of Computer Vision Problems in Plant Phenotyping (CVPPP) 2014 and aims at seeking for fully automated approaches. For the challenges the organization release a training sets (containing raw images and annotations) and testing sets (containing raw images, only). The right two columns in Fig. 1.12 show samples of raw images and annotations. The raw images were captured with a top-view camera. Although the leaf color is different from background (soil) color, the segmentation of leaf is not easy to implement due to the occlusion of leaves by other leaves. Because the fact that the color of a leaf might be changed with lights or different camera parameters, any approach, even if it performs well on this dataset of challenges, will probably not transfer well to other datasets.

Moreover, when the *Arabidopsis* grows quickly and occlusions becomes more heavy, any approach of segmentation would be impractical.


Figure 1.12: Sample images of the Leaf Segmentation and Counting Challenges. Left: image acquisition and plants growing environment. Right: two columns show samples of raw images and annotations respectively. Images are from CVPPP2014.

Another key issue of the 2-D imaging techniques is the calibration of camera. In most modern phenotyping platforms, the camera needs to be calibrated in advance with manual or semi-automatic manner. For example, although LemnaTec platform is capable of automated capturing plants using advanced digital cameras at multiple viewpoint, it is still difficult to capture the various plants with optimal field of view or focal length. Therefore, the quality of captured image is not always satisfied for diverse plants with different shapes. Fig. 1.13 illustrates the limitation: while with high-resolution pixels, the effective section occupied by the maize is only approximately half area of the image, the maize

lacks the detailed textures, so that image feature detection could fail in this case.

Figure 1.13: A raw image (2448×2050 pixels) of a maize plant, which captured by LemnaTec platform. The effective section (red rectangle) occupied by the maize is only approximately quarter of the image size. Cropped image (1224×1025 pixels) lacks the detailed surface textures.

### 1.2.2 Motivation

As mentioned above, phenotyping involves the objective measurement of characteristics or traits. 3-D imaging technologies (3-D laser scanner, structured light sensor, etc.) are increasingly incorporated into mass produced consumer goods and have the potential applications providing a cost-effective alternative to current commercial phenotyping platforms.

For example, Li *et al.* (Li *et al.*, 2015) developed a 4-step approach that can automatically detect and segment the stems of tomato plants using a Kinect sensor (Kinect, 2011), including acquisition and pre-processing of image data,

detection of stem segments, removing false detections and automatic segmentation of stem segments. The results are shown in Fig. 1.14, where: (a) is an original color image obtained from the Kinect; (b) is the depth image obtained from the Kinect; (c) is the color image after aligning (a) and (b); (d) is the binary image after thinning of (b); In (e), all detected stems are labeled in the depth image; In (f), the central axis is colored red, and the grid system is constructed by gray lines.

It is clear that the 3-D model (Fig. 1.14(c) ) yielded by Kinect sensor missed many tiny or thin parts, such as small branches and leaves. Experimental results in (Paulus *et al.*, 2014c; Khoshelham & Elberink, 2012) show that with increasing distance, the accuracy of Kinect depth measurement decreases from a standard deviation (SD) of a few millimeters to about 40 mm, and the point-to-point distance increases from 0.9 mm to 7mm − this inherent limitation explains why Kinect sensor only works on the plants with wide leaves and always fails on detecting the main stems and branches. In our experiments, Kinect sensor cannot detect out the stems or leaves whose size are smaller than 0.5 cm.

Paulus *et al.*(Paulus *et al.*, 2014a) used a 3-D scanner system (∼£70,000), which consists of a line laser scanner (*Perceptron V5 scanner*) and a 3-D articulated measuring arm (*Romer Infinite 2.0*), to obtain 3-D point clouds of two barley plants in time course, and the organs (leaf, stem) were then detected with classification algorithms. The scanning experiments are shown in Fig. 1.15. The accuracy of the 3-D scanner system is high, however, it is very expensive and has to use the commercial post-processing software (registration, filtering, integration) to obtain the final 3-D model.

Although commencing as a relatively simple structure, the plant body rapidly becomes more complex due to re-iterative organ formation, changes in organ spacing and identity, and branching that leads to overlapping and variable 3-D organization. Fig. 1.16 shows the complexity of structure changing of a *Arabidopsis* growing at different stages.

Recording this complexity in a dynamic (non destructive) manner that retains accurate 3-D spatial information in a retrievable format remains a serious challenge in biology.

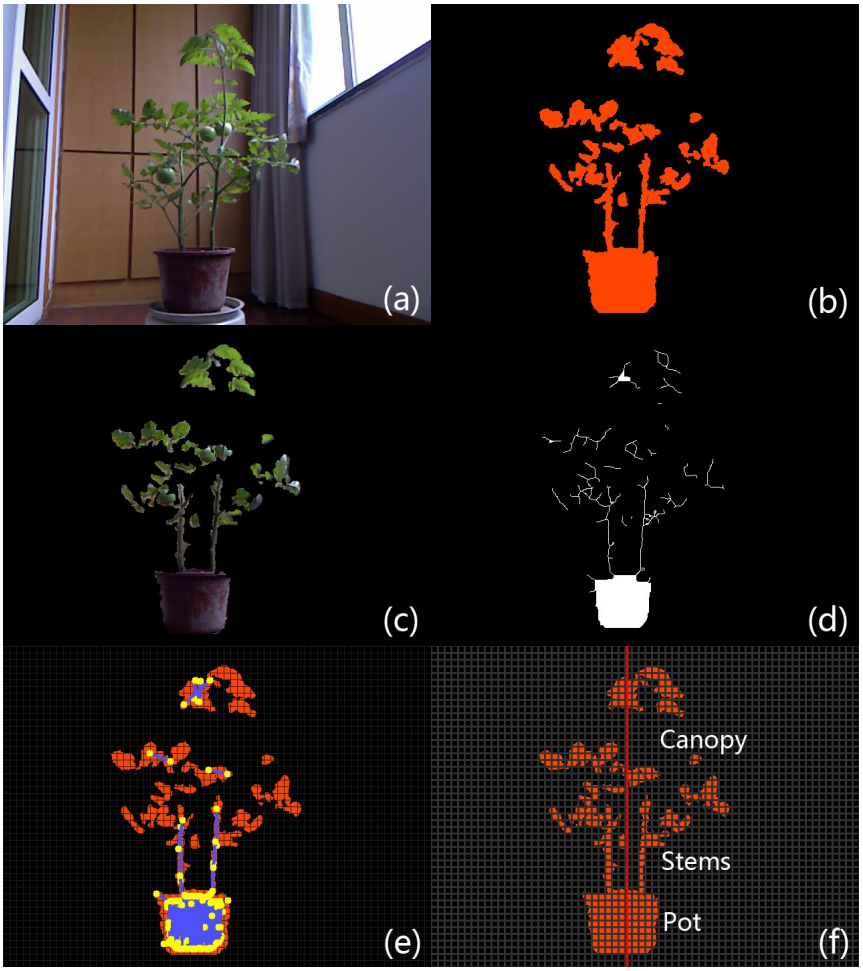



Figure 1.14: Detection and segmentation of the stems of tomato plants using a Kinect sensor. (a) is an original color image obtained from the Kinect; (b) is the depth image obtained from the Kinect; (c) is the color image after aligning (a) and (b); (d) is the binary image after thinning of (b); In (e), all detected stems are labeled in the depth image; In (f), the central axis is colored red, and the grid system is constructed by gray lines. Image courtesy of (Li *et al.*, 2015).

The development of 3-D sensor technologies has been traditionally driven by the 3-D game, 3-D CAD modeling (reverse engineering) and recently by 3-D printing, rather than by 3- D modeling of plants. While the majority of existing 3-D imaging techniques or devices are able to generate depth images or point clouds for digitization of plants, so far these techniques are usually expensive or impractical for routine measurement and analysis, especially for large-scale implementation in a greenhouse (Paulus *et al.*, 2013; Paulus *et al.*,

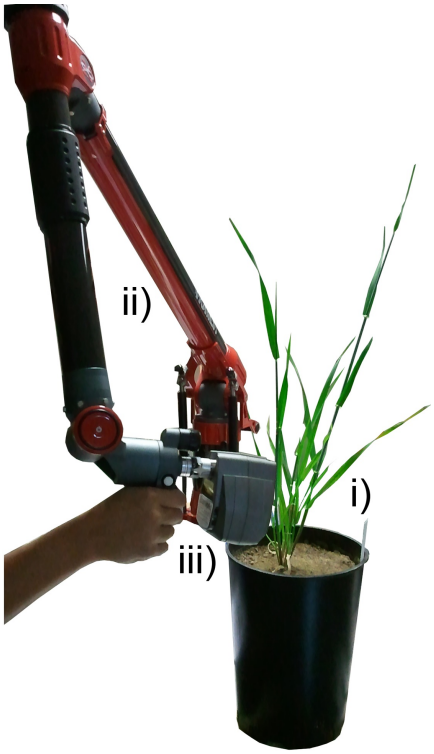



Figure 1.15: Laser scanning of a barley plant. (i) measuring arm (ii) coupled device (iii) a measuring sphere of 1.4 m in radius (Romer Infinite 2.0 and Perceptron V5, Plymouth, MI, USA). Image courtesy of (Paulus *et al.*, 2014a).

2014a; Vos *et al.*, 2010; El-Omari & Moselhi, 2011). Furthermore, many of these techniques either depend on complicated manual procedures (e.g. camera calibration, interactively post-processing), or are case-specific in application, which means the system must be rearranged significantly for new circumstances or experiments.

Apart from high-cost and inconvenience mentioned above, available 3-D solutions have the following limitations:

(1) Limited to specific organs (e.g. leaf (Kaminuma *et al.*, 2004) (Biskup *et al.*, 2007) (Konishi *et al.*, 2009) (Apelt *et al.*, 2015) (Pound *et al.*, 2014) (Li *et al.*, 2015));

(2) Only applicable to a single plant (e.g. *Arabidopsis* (Kaminuma *et al.*, 2004), cotton (Paproki *et al.*, 2012), maize (Ivanov *et al.*, 1995), cereal (Paulus *et al.*, 2014b), tomato (Li *et al.*, 2015).);

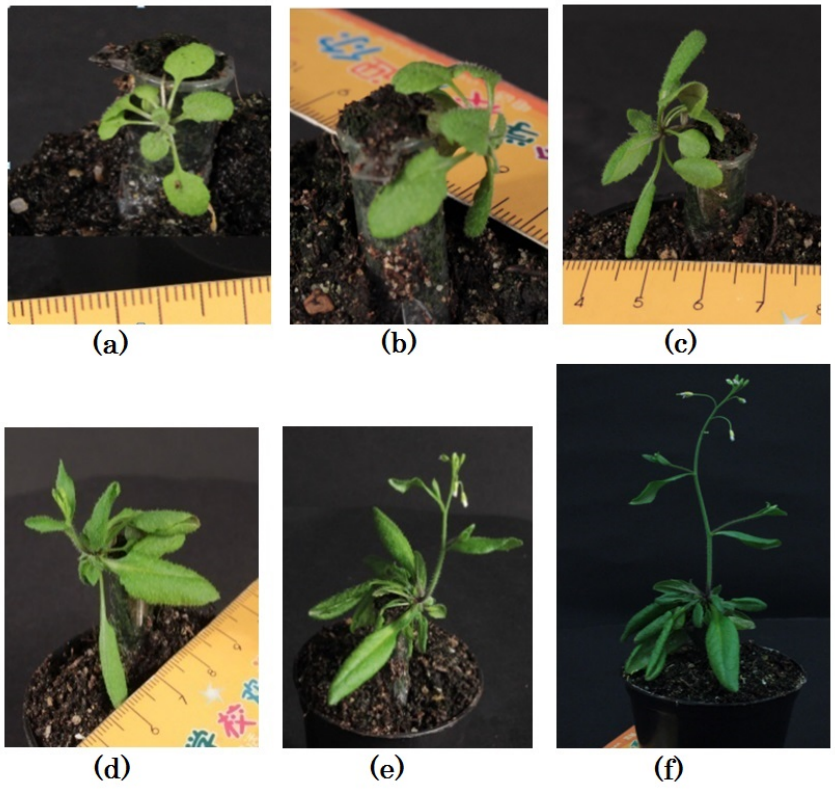



Figure 1.16: The changeable structures of a *Arabidopsis* growing at different stages. (a)∼(f) indicate the *Arabidopsis* sample images took at 18, 21, 24, 29, 32 and 35 days after seeding, respectively.

(3) Only suitable for early growing stages when architectures are relatively simple without significant occlusions (Li *et al.*, 2015).

Therefore, accurate full 3-D geometrical plant models, especially those able to record various plants species during the whole growth period, would be attractive for phenotyping. Such models can contain the information needed to compute a variety of plant traits, such as total leaf shape and area, branch angle, and are essential where other data (i.e. fluorescence, thermal) needs to be co-registered with the complex surface morphology of the plant.

Therefore, there is an urgent need for rapid, automated, and generalized techniques for accurately reconstructing and measuring the 3-D architecture of complex plants. This need has been an inspiration to my research.

## 1.3 Research Aim and Objective

The aim of my research was to design and develop an accurate 3-D reconstruction framework through the use of low-cost digital photography equipment and computer vision techniques, to guide the actions of 3-D imaging based resolution for plant phenotyping.

The objectives of this research can be summarised as follows:

(1) Design and develop a low-cost practical multi-view 2-D image acquisition system to resolve the plant self-occlusion problem as much as possible, and conduct data acquisition experiments to validate the effectiveness of the system.

(2) Develop camera parameters and pose estimation methods from the captured uncalibrated multi-view images of plants, and validate the performance of this method through conducting experiments.

(3) Develop a 3-D reconstruction method based on multi-view images to yield dense, complete and detailed 3-D point cloud, and validate the accuracy and completeness of 3-D reconstruction results by comparison analysis.

(4) Develop a 3-D measurement approach to extract geometric features directly from resulting 3-D reconstruction results, and perform experiments to validate the effectiveness of the framework.

## 1.4 Structure of Thesis

The remaining chapters of this thesis are organised as follows:

- Chapter 2, the background of existing 2-D/3-D imaging methods for plant phenotyping is provided, including 2-D digitalizing and visualization, 3-D sensor techniques and a brief introduction on digital camera.

- Chapter 3, this chapter presents a low-cost multi-view 2-D image acquisition system to deal with the plant's complex shape and self-occlusion problem, and introduces some typical plant phenotyping experiments.

- Chapter 4, this chapter describes a camera parameter estimation method from uncalibrated multi-view images, which include: i) extraction of local invariant features for each image; ii) feature matching from neighboring overlapped image pairs, and iii) estimation of camera parameters and pose based on incremental Structure from Motion. The experimental results are discussed.

- Chapter 5, an accurate multi-view stereo 3-D reconstruction method was proposed to generate more dense, detailed and complete 3-D point cloud, which includes stereo pair selection, depth maps computation, depth maps refinement and depth maps merging. The 3-D reconstruction results are evaluated through comparison analysis.

- Chapter 6, this chapter presents a novel automatic 3-D point cloud segmentation approach based on spectral clustering algorithm, as well as an interactive automatic method on estimation of branch angle using a 3-D mini-bounding box approach.

- Chapter 7, A discussion of the achievements of this research is presented in the context of the current state-of-the-art and future research directions are proposed.

# Chapter 2

# Background

In recent years, many computer vision techniques are applied to plant structure analysis and phenotyping. This chapter focuses on pre-existing research in the area of 3-D imaging techniques for plant structure analysis or phenotyping. Then, a number of concepts and principals of digital camera and camera model in computer vision are described as necessary to set a background for the present work.

## 2.1 Imaging Based Plant Phenotyping

Over the past decades, different imaging sensors, image processing techniques, robotics, and high performance computing have been applied to plant phenotyping.

### 2.1.1 The Outline of 2-D Imaging Methods

As mentioned above, 2-D imaging and analysis techniques have been studied and applied to many tasks, such as crop detection and segmentation, leaf and fruit recognition, plant disease and pest analysis. Various automated high-throughput platforms for plant phenotyping have emerged in recent years with the goal of producing large amounts of phenotype data, including PHENOPSIS (Granier *et al.*, 2006) , TraitMill$^{TM}$ , LemnaTec platforms (LemnaTec., 2014).

To review the state of the art of phenotyping platforms, we present a list of recent publications (Table.2.1). The platform of LemnaTec was used in about

a half of these. Bespoke platforms were used in other publication but these are rarely automated. This illustrates that integrative automated high-throughput phenotyping measurements/platforms are still rather rare (Humplík *et al.*, 2015; Li *et al.*, 2014b). In addition, most of phenotyping measurements/platforms have common limitations as follows:

(1) Focus on high-throughput measurement, but they pay less attentions on accuracy or completeness of the results.

(2) Often only applicable to a single plant, mostly *Arabidopsis*, or a few types of related plants.

(3) About 50% were applied cases rather than methodology development.

The above mentioned 2-D imaging techniques can provide projected images of objects instead of complete 3-D plant structural information. Therefore, they can only obtain relatively simple shape and texture properties (LemnaTec., 2014). These inherent limitations make 2-D imaging based phenotyping systems less accurate, especially for the plants with complicated and self-occlusion structures.

| Study | Plant species | Type of the study | Type of analysis | Platform/origin |
|---|---|---|---|---|
| Humplk et al.(2015) | pea, field cultivars | methodology | RGB (multi-views) | PlantScreen |
| Clauw et al.(2015) | Arabidopsis | applied | RGB (top-view) | WIWAM |
| Neumann et al.(2015) | barley | methodology | RGB (multi-views) | LemnaTec |
| Neilson et al.(2015) | Sorghum | methodology | RGB (multi-views ), NIR | LemnaTec |
| Chen et al.(2014) | barley | methodology | RGB (multi-views), NIR, | LemnaTec |
| Petrozza et al.(2014) | tomato | applied | RGB (multi-views), NIR, | LemnaTec |
| Harshavardhan et al.(2014) | Arabidopsis | applied | RGB (top-view), NIR | LemnaTec |
| Honsdorf et al.(2014) | barley,(wild species) | applied | RGB (multi-views) | LemnaTec |
| Coupel-Ledru et al.(2014) | grapevine | applied | RGB (multi-views) | LemnaTec |
| Poiret al.(2014) | Brachypodium | methodology | RGB (multi-views ) | LemnaTec |
| Schilling et al.(2014) | barley | applied | RGB (multi-views) | LemnaTec |
| Hairmansis et al.(2014) | rice | applied | RGB (multi-views) | LemnaTec |
| Vasseur et al.(2014) | Arabidopsis | applied | RGB (top-view) | PHENOPSIS |
| Bresson et al.(2013) | Arabidopsis | applied | RGB (top-view) | PHENOPSIS |
| Tisnet al.(2013) | Arabidopsis | applied | RGB (top-view) | PHENOSCOPE |
| Cseri et al.(2013 | barley | methodology | RGB (multi-views) | self-construction |
| Pereyra-Irujo et al.(2012) | soybean | methodology | RGB (two-views) | self-construction |
| Skirycz et al.(2011) | Arabidopsis | applied | RGB (top-view) | WIWAM |
| Golzarian et al. (2011) | barley | methodology | RGB (multi-views) | LemnaTec |
| Harris et al.(2010) | wheat, barley | applied | RGB (multi-views) | LemnaTec |
| Jansen et al.(2009) | Arabidopsis, tobacco | methodology | RGB (top-view) | self-construction |
| Rajendran et al.(2009) | wheat | applied | RGB (multi-views) | LemnaTec |
| Chaerle et al.(2007) | bean | methodology | RGB (top-view) | self-construction |
| Chaerle et al.(2006) | tobacco | methodology | thermoimaging, | self-construction |
| Granier et al.(2006) | Arabidopsis | methodology | RGB (top-view) | PHENOPSIS |

Table 2.1: List of selected works describing 2-D imaging analysis.

### 2.1.2 3-D Imaging Methods

3-D imaging of plants refers to the construction of 3-D models that allow measurement, analysis and simulation based on the specimens' geometric features (Santos & Oliveira, 2012).

Advanced 3-D imaging techniques are key to the realization of accurate 3-D model reconstruction for plant structure analysis. Currently, there are 3-D imaging techniques (light detection and ranging (Lidar), Time-of-Flight (ToF) camera, 3-D light field camera, CT scanner, Structured light sensor and stereo/multi-view vision) that provide more or less accurate models:

(1) Laser Detection and Ranging (Lidar).

Light detection and ranging (Lidar) is a remote sensing technology that measures distance by illuminating a target with a laser and analyzing the reflected light. It is not susceptible to ambient sunlight and can work reliably outdoors. Some researchers have positioned the Lidar several meters away from relatively large-scale sensing objects to estimate the height (Hopkinson *et al.*, 2004; Hosoi *et al.*, 2005) and canopy structure (Hosoi & Omasa, 2006; Zande *et al.*, 2007) of trees and the carbon stock of forest. Kaminumaet *et al.*, (Kaminuma *et al.*, 2004) applied a laser range finder to reconstruct 3-D models that represented the leaves and petioles as polygonal meshes and then quantified morphological traits from these models. The structural parameters of trees and forest, such as tree height, stem count, biomass, material distribution, etc., are provided by these systems. The dry weight of the stem, leaves, and ears of a corn field could also be estimated based on the density distribution of 3-D points captured by Lidar sensors (Hosoi & Omasa, 2009). The recent Lidar technology development has enabled a high-resolution 3-D sensor based on the optical-probe method with a range accuracy of 0.5mm at a distance of 3.5m. The high-density and clean 3-D data of aubergine and sunflowers were generated by a Lidar in (Omasa *et al.*, 2007), and the author demonstrated the 3-D point cloud composited with textures, such as natural color, chlorophyll fluorescence, and leaf temperature information.

One limitation of Lidar system is that it does not capture the 3-D points

of one frame simultaneously, and therefore, it requires that the sensor and the plant be kept relatively still during the process, making the system difficult to deploy on a mobile platform or when conducting in-field phenotyping tasks in windy day (Li, 2014).

(2) Time-of-Flight (ToF) Camera.

A ToF camera is a range imaging camera system that measures distance based on the known speed of light, measuring the time-of-flight of a light signal between the camera and the object in the field of view. Two popular ToF cameras, PMD Camcube (PDM Tec, Germany) and IFM Efector 3-D (IFM, USA), were reported to have a depth resolution (∼1cm) and to be usable with an autonomous phenotyping robot when no heavy dust exists (Klose *et al.*, 2009). 3-D plant modeling was reported by combining a 2-D color camera and a 3-D ToF camera, whose position and attitude was controlled and measured using a robot arm (Alenyà *et al.*, 2011). In these methods, authors combined 3-D spatial data with 2-D color information and performed segmentation based on color in 2-D space, instead of segmenting the plant from the background in 3-D space, which relied on a fused depth/color image. If the robotic arm is of high positioning accuracy and precision, the approach may get high accuracy and reliability but at considerable cost. On the contrary, (Rusu *et al.*, 2008) developed a low-cost robust multiple-viewpoint cloud alignment approach based on 3-D feature histograms, which is purely based on 3-D image processing without the aid of extra devices (e.g. high-precision robotic arm).

Fig. 2.1 illustrates that a ToF 3-D laser was used to scan an apple tree from 2∼3 different viewpoints respectively, and generate a 3-D point cloud by post-processing software.

(3) 3-D light field camera.

A 3-D light field camera is another newly available 3-D sensor (GmbH, 2014). It has a micro-lens array, and each micro-lens captures a tiny image. Therefore, the resulting raw image is a composition of tiny images. Software is used to extract the depth information based on the correspondence of these images. The authors (Apelt *et al.*, 2015) used the 3-D light

Figure 2.1: A ToF 3-D laser (left) was used to scan an apple tree from 2∼3 different viewpoints, and then a 3-D point cloud (right) of the tree was produced by post-processing − registration, filtering and integration. Image courtesy of INRA.

field camera to monitor rosette and individual leaf growth in *Arabidopsis* in 4-D with high spatio-temporal resolution.

This technique features high spatial resolution and accuracy and can work reliably outdoors under sunlight. However, its price is significantly higher than the ToF camera. (Apelt *et al.*, 2015)

(4) CT scanner. Industrial CT (computer tomography) uses a series of 2-D images taken at specific intervals around the entire sample. Once the acquisition process of the CT scan is completed, CT calibration and CT reconstruction algorithms are used to reconstruct the 3-D CT volume. These 3-D images are made of voxels. The main benefit of using 3-D CT for scanning or digitization is that a complete model can be obtained with both external and internal surfaces of an object without destroying it. Some researchers used the CT scanner to analysis tillers and grain quality of rice (Yang *et al.*, 2011) and wheat (Karunakaran *et al.*, 2004; Garbout *et al.*, 2012; Gregory *et al.*, 2003; Flavel *et al.*, 2012). However, its price is high.

(5) Structured light sensor. Structured light sensor projects structured light,

such as parallel lines and grids, on to the object surface. The structured light is distorted by the object surface, and therefore the 3-D shape of the object can be recovered by analyzing the distorted pattern. The author (Veeken, 2006) reported a corn plant sensing system that estimated the 3-D structure of a corn plant by projecting parallel lines to the plant. When structured light sensor applied to a plant, the major challenge is the mismatch or unavailability of correspondence caused by texture-less leaves, complex structure, and occlusion problems, that could lead to the failure of 3-D reconstruction. Additionally, changing lighting conditions, which is common for the plants grown in the field outdoors, could influence the performance of 3-D reconstruction (Omasa *et al.*, 2007; Weiss *et al.*, 2010).

Structured light sensor has been used to achieve great success for indoor applications recently. The David structured light scanner (David Structured Light Scanner SLS-2, David Vision System GmbH, Germany) can generate high-quality 3-D images with precision up to 0.1% of the scan size (down to 0.06 mm) (Paulus *et al.*, 2014c). However, these methods are vulnerable to strong ambient light, and work poorly outdoors except in the evening or at night.

The Kinect sensor, at a very affordable price ranging from $99 to $200, originally designed as a motion sensing input device by Microsoft for the Xbox game console, is a physical device with depth sensing technology, a built-in color camera (1080p, 30Hz ), an infrared (IR) emitter, and a microphone array, enabling it to sense the location and movements of people as well as their voices (depth sensing resolution: 640×480, 30Hz, FOV: 70×60, range distance: 0.5∼4.5 meters) (Kinect, 2011). In 2011, Microsoft released the first version Software Development Kit (SDK) for Kinect, which enables users to develop sophisticated computer-based human body tracking applications. As a low-cost, portable, non-intrusive and novel detecting sensor, Kinect sensor has attracted many researchers to pay great attention to applying the Kinect technology for different tasks, including object tracking, facial recognition, human activity analysis, hand gesture recognition, robot indoor 3-D mapping, etc. (Han *et al.*,

2013).

Li *et al.* developed a 4-step low-cost structured light sensor approach (Li *et al.*, 2015) that can detect and segment the stems of tomato plants using Kinect sensor, including acquisition and pre-processing of image data, detection of stem segments, removing false detections and automatic segmentation of stem segments. Li *et al.* (Li *et al.*, 2013) presented a framework to track and detect plant growth by a forward-backward 3-D point cloud analysis, where the 3-D point cloud was produced based on a structured light scanner over time. The Kinect (Kinect, 2011), David laser scanner (Winkelbach *et al.*, 2006), and Perceptron V5 scanner system were analyzed and compared in the research on plant phenotyping (Paulus *et al.*, 2014c)(Li *et al.*, 2015). However, when applied for plants, the 3-D models yielded by Kinect sensor usually miss many tiny or thin parts, such as small branches and leaves. Fig. 2.2 and Fig. 2.3 show two samples of 3-D modeling using structured light and Kinect sensor respectively.

(6) Stereo/multi-view vision. Stereo/multi-view vision senses the same object by two or more 2-D cameras that are placed at distinct locations, and they estimate the depth information by searching for the correspondence of the image pairs. Researchers have applied this technique to analyze the transplant variables of sweet potatoes (He *et al.*, 2003) , and to reconstruct a 3-D model of corn canopy for phenotyping (Ivanov *et al.*, 1994; Ivanov *et al.*, 1995). Biskup designed a stereo vision system with two cameras to build 3-D models of soybean plant foliage to analyze the angle of inclination of the leaves and its movement over time (Biskup *et al.*, 2007). Quan proposed a method to interactively create a 3-D model of the foliage by combining clustering, image segmentation and polygonal models (Quan *et al.*, 2006). 3-D plant analysis based on a mesh processing technique was presented in (Paproki *et al.*, 2012), where the authors created a 3-D model of cotton plant from high-resolution images using a commercial 3-D digitization product named 3DSOM. Thiago developed an image-based 3-D digitizing method for plant architecture analysis and phenotyping (Santos & Oliveira, 2012), and showed that the state-

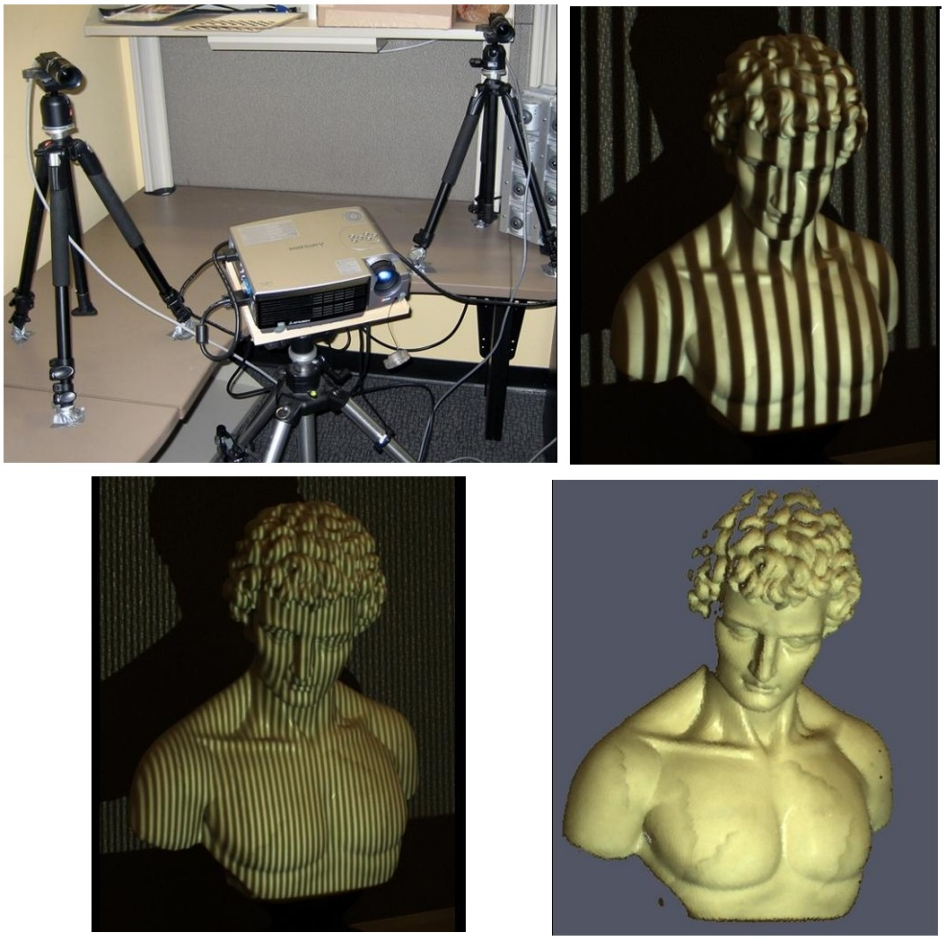


Figure 2.2: A statue (left) and its 3-D model (right) produced by a structured light sensor. Image courtesy of (Lanman & Taubin, 2009).

> of-the-art SFM and multi-view stereovision are able to produce accurate 3-D models with a few limitations. Yamazaki *et al.* presented a practical shape-from-silhouettes approach to acquire 3-D models of intricate objects with severe self-occlusions, repeated thin structures, and surface discontinuities, including tree branches, bicycles and insects (Yamazaki *et al.*, 2009).

Although the development of advanced 3-D sensors makes 3-D spatial data available with relatively reasonable resolution, accuracy, and noise, 3-D reconstruction algorithms for plant phenotyping are still primitive. These methods for 3-D reconstruction of plants does not achieve adequate accuracy with reasonable processing speed or a more affordable system cost (Li, 2014). More

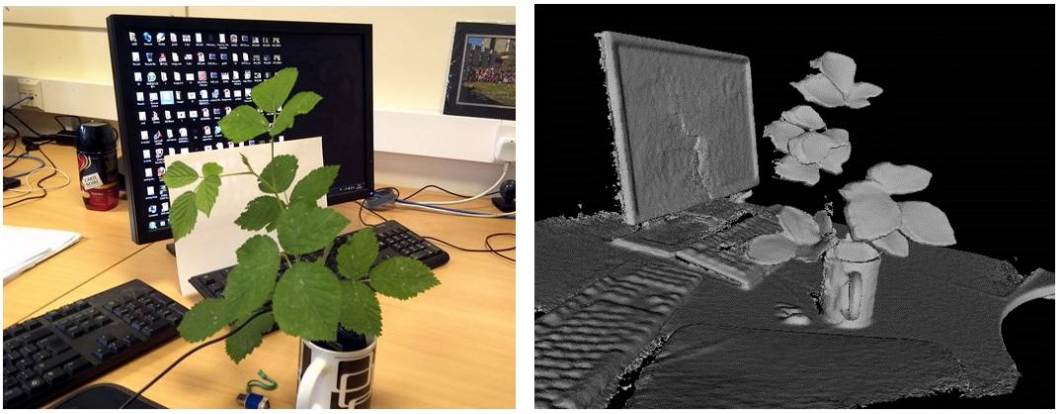


Figure 2.3: A desktop scene (left) and its 3-D models (right) produced by a Kinect sensor and software *ReconstructMe*(ReconstructMe, 2014). The 3-D plant model missed small stems and leaves.

efforts still need to be made to perform 3-D reconstruction and extract detailed geometric features with the accuracy, processing speed, system cost and reliability desired by plant biologists.

## 2.2 Introduction to the Digital Camera

Nowadays, the digital cameras are widely used in the world. Point the lens at your subject, press the shutter button, and you have captured an image that ultimately produces a photograph. Among of them, the DSLR (digital single lens reflex) cameras are hugely popular cameras with amateur and professional photographers, which use a prism and mirror based system that allows the photographer to see exactly what they are taking (see Fig. 2.4). The sensors in most DSLR cameras are made of a regular matrix of tiny devices that measure the intensity of light falling onto them, The common sensor type is CCD (charge-coupled device), another sensor type is CMOS (Complementary Metal-oxide-semiconductor) that used in some newer camera products. The technology within the sensor and the quality of light refracted by the lens will directly affect the quality of the final image.

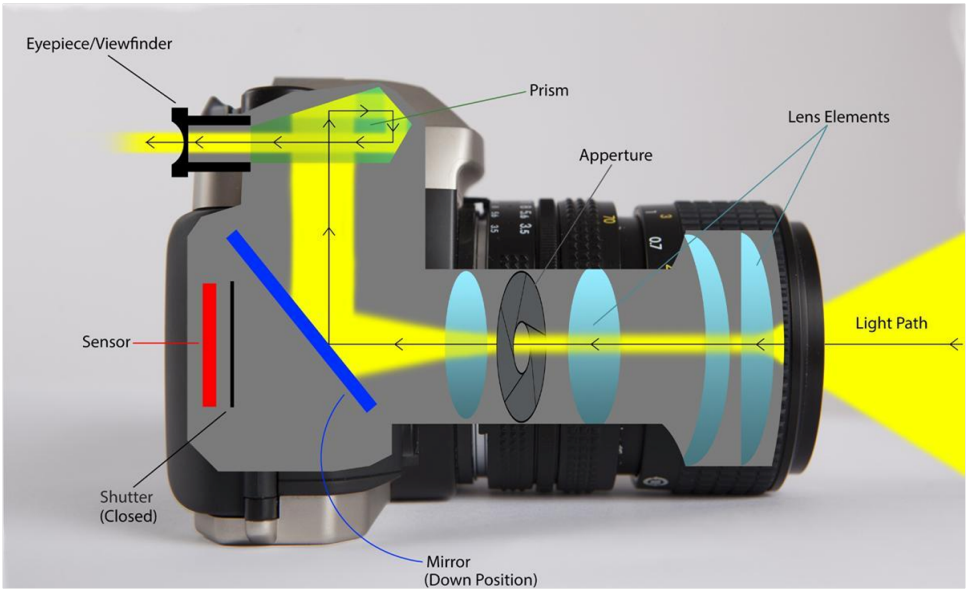



Figure 2.4: Structure of a DLSR camera. Image courtesy of Panasonic.

### 2.2.1 Pinhole Camera Model

In computer vision, a camera could be approximated by an ideal projective model, often called *pinhole model* (Faugeras, 1993). The pinhole camera is described by its optical center $C$ (camera projection center) and the image plane, See Fig. 2.5. The distance of the image plane from $C$ is the focal length $f$. The plane parallel to the image plane containing the optical center is called the principal plane or focal plane of the camera. If $(X, Y, Z)$ is the coordinates of a 3-D point $M$, $(u, v)$ is the image coordinates of the projection 2-D point $m$ of $M$, then the relationship between image coordinates and 3-D space coordinates can be given by

$$\frac{u}{X} = \frac{v}{Y} = \frac{f}{Z}, \tag{2.1}$$

which can be written in a linear framework as

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \tag{2.2}$$

However, if we want to deal with issues such as focus, exposure, vignetting,

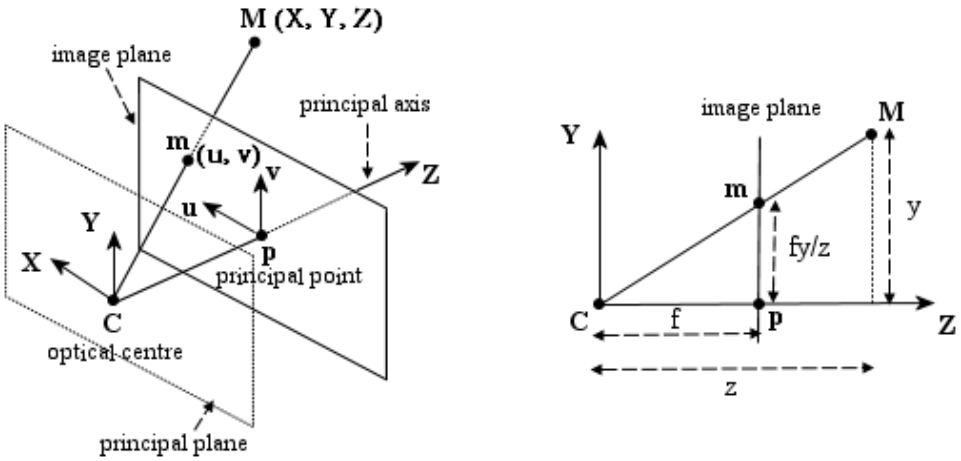



Figure 2.5: Pinhole model of camera. The left figure illustrates the projection of the point $M$ on the image plane by drawing the line through the camera centre $C$ and the point to be projected. The right figure illustrates the same situation in the $YZ$ plane, showing the similar triangles used to compute the position of the projected point $m$ in the image plane.

and aberration, we need to use a more sophisticated model – *thin lens model*, shown as Fig. 2.6. According to the *lens law* (Carl, 2014), the relationship between the distance to an object $v$ and the distance behind the lens at which a focused image is formed $b$ can be expressed as:

$$\frac{1}{v} + \frac{1}{b} = \frac{1}{f}, \tag{2.3}$$

where $f$ is called the focal length of the lens. The angle of view ($2\alpha$) depends on the ratio between the sensor width $W$ and the focal length $f$ (or, more precisely, the focusing distance $b$, which is usually quite close to $f$ in a real camea) (Szeliski, 2010).

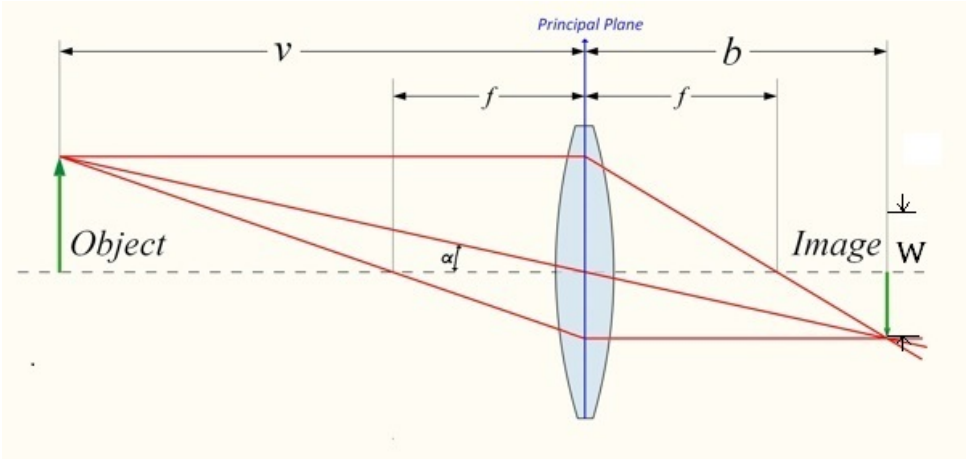



Figure 2.6: Thin lens model of a camera. The relationship between the distance to an object $v$ and the distance behind the lens at which a focused image is formed $b$ can be expressed as: $\frac{1}{v} + \frac{1}{b} = \frac{1}{f}$, where $f$ is called the focal length of the lens. Image courtesy of (Jongerius, 2012).

### 2.2.2 Camera Geometry

In camera model, a point $M(X, Y, Z)$ in 3-D space is projected to the 2-D point $x(u, v)$ of image plane with the following formula

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{2.4}$$

which can be represented as

$$x = \mathbf{P}M, \tag{2.5}$$

where $\mathbf{P}$ is a $3 \times 4$ matrix called as the *projection matrix* (Faugeras, 1993) which models the pinhole camera.

The above projection equations are derived under the assumption that the camera is situated at the origin of the world coordinate system, with its principal axis pointing along the positive Z-axis. In general, when a real and lens-based camera is moved in the real world coordinate system, its projection matrix

$\mathbf{P}$ can be written as a special structural given by (Hartley & Zisserman, 2004)

$$\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}], \tag{2.6}$$

where matrix $\mathbf{K}$ is called as *intrinsic matrix* that encodes the internal parameters of the camera, and is parameterized by Hartley and Zisserman (Hartley & Zisserman, 2004) as

$$\mathbf{K} = \begin{bmatrix} f_x & s & c_u \\ 0 & f_y & c_v \\ 0 & 0 & 1 \end{bmatrix}. \tag{2.7}$$

The intrinsic matrix $\mathbf{K}$ describes internal geometric parameters of the camera as the follows:

- **Focal length**. The focal length ($f_x$,$f_y$) is the distance between the camera center and the image plane, which is measured in pixels. In common cases, both $f_x$ and $f_y$ have the same value $f$.
- **Principal point offset**. The camera's principal axis is the line perpendicular to the image plane that passes through the pinhole. Its itersection with the image plane is referred to as the principal point. The principal point offset $(c_u, c_v)$ is the location of the principal point relative to the origin (usually is defined at the top left corner) of the image plane.
- **Axis skew**. Axis skew $s$ is only used if the pixel array in the sensor is skewed and can safely be set to 0 in most cases.

The projection matrix $\mathbf{P}$ specifies the transformation from world to camera coordinates, which is a standard 3D coordinate transformation, including the rotation matrix $\mathbf{R}$ and the translation vector $\mathbf{t}$.

We can summarize the projection matrix of the camera below:

$$\mathbf{P} = \overbrace{\mathbf{K}}^{\text{Intrinsic Matrix}} \times \overbrace{[\mathbf{R} \mid \mathbf{t}]}^{\text{Extrinsic Matrix}}.$$

### 2.2.3 EXIF and Intrinsic Matrix

When the user takes a photo with a digital camera (see Fig. 2.7), the EXIF (Exchangeable Image File Format, (Photometadata, 2015) ) tag of the output JPEG file often includes information about camera maker, camera model, image width and height, exposure time, focal length, aperture, ISO degree, flash mode and taken date, etc. Some of them are very useful for further processing. For example, camera model can be used for finding out the sensor size of the camera from some digital camera databases, e.g. (Database, 2015) or (Dpreview, 2015). Hence, we can gather such useful information to calculate intrinsic matrix.

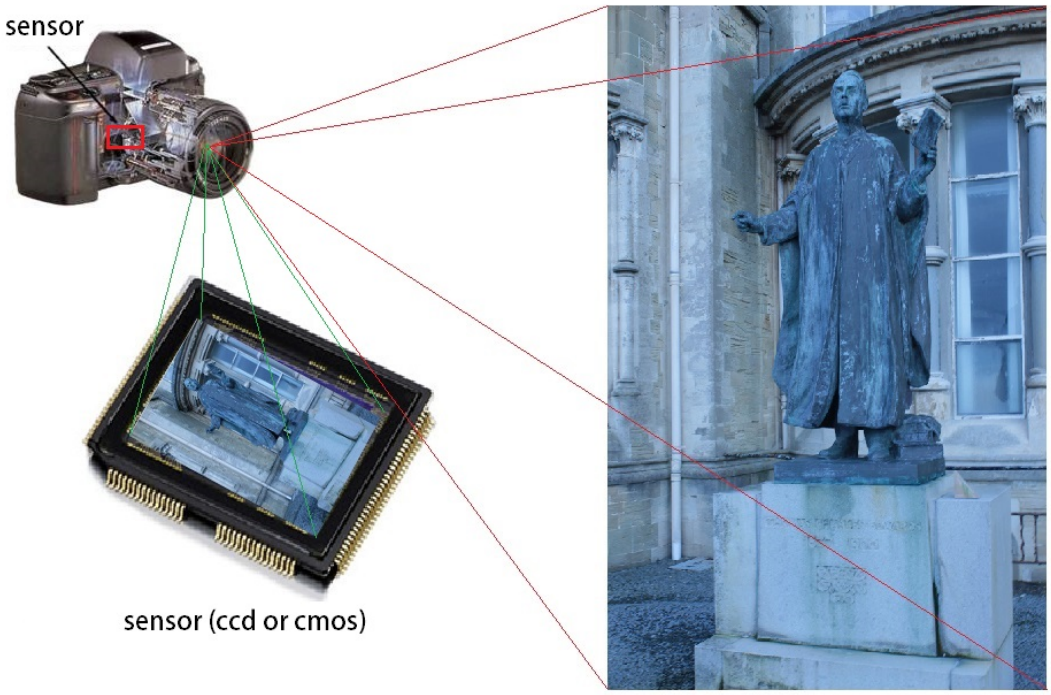



Figure 2.7: Take a picture - from world units to pixels.

Among these parameters, the focal length $F_{mm}$ is especially useful for camera recovery (described in the following chapters). Usually, the focal length $F_{mm}$ is expressed in millimeters.

In order to calculate the intrinsic matrix, we need to extract the focal length from the EXIF tag of a JPEG file and convert it to pixel units, so that it can be used in the camera recovery. First, we can get the CCD sensor's width $W_{CCD}$ from digital camera database according to the camera model. Then, the focal

length $f_i$ (in pixels) can be computed from the focal length $F_{mm}$ (in $mm$), the image width $W_i$ (in pixels) and the CCD sensor's width $W_{CCD}$ (in $mm$) by the following formula:

$$f_i = \frac{W_i \times F_{mm}}{W_{CCD}}. \tag{2.8}$$

Therefore, intrinsic matrix $\mathbf{K}$ is calculated as:

$$\mathbf{K} = \begin{bmatrix} f_i & 0 & c_u \\ 0 & f_i & c_v \\ 0 & 0 & 1 \end{bmatrix} \tag{2.9}$$

where $c_u$ is the half of image width, and $c_v$ is the half of image height.

### 2.2.4 Deviations From Camera Model

The pinhole model is only an approximation of the real camera projection. Modern camera lenses are relatively free of geometric distortion. However, there is always a small remaining amount even with the most expensive lenses. Radial distortion is a linear displacement of image points radially to or from the center of the image, caused by the fact that objects at different angular distance from the lens axis undergo different magnifications (Pollefeys, 2002). Radial distortion is negligible under most circumstances, however, when high accuracy is required, or special lenses (i.e. short-focal, fish-eye) are used, radial distortion has to be taken into account (Wu, 2014).

Fig. 2.8 shows different forms of radial distortion.

Fortunately, compensating for radial distortion is not that difficult in practice. For most lenses, a simple quartic model of distortion can produce good results (Szeliski, 2010). The coordinates in undistorted image plane coordinates $(x, y)$ can be obtained from the observed image coordinates $(x_o, y_o)$ by the following equation (Hartley & Zisserman, 2004):

$$\begin{aligned} x &= x_o + (x_o - c_x)(K_1 r^2 + K_2 r^4 + ...) \\ y &= y_o + (y_o - c_y)(K_1 r^2 + K_2 r^4 + ...) \end{aligned} \tag{2.10}$$

where $K_1$ and $K_2$ are the first and second *radial distortion parameters* of the

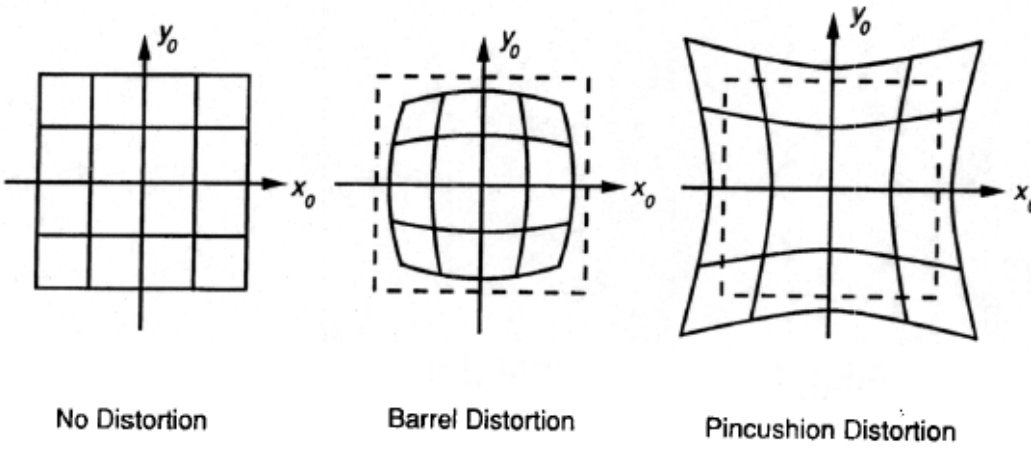



Figure 2.8: Radial distortion of lens. Image courtesy of (Ann Arbor, 2014).

radial distortion, and

$$r^2 = (x_o - c_x)^2 + (y_o - c_y)^2 \tag{2.11}$$

The radial distortion parameters of a lens ($K_1$, $K_2$) are usually calculated with pre-calibration, such as (Zhang, 2000). However, when the *focal length* of the camera changes (through zoom or focus), the parameters $K_1$ and $K_2$ will also vary and that approximation can be modeled as follows (Pollefeys, 2002):

$$\begin{aligned} x &= x_o + (x_o - c_x)(K_{f1}\frac{r^2}{f^2} + K_{f2}\frac{r^4}{f^4} + ...), \\ y &= y_o + (y_o - c_y)(K_{f1}\frac{r^2}{f^2} + K_{f2}\frac{r^4}{f^4} + ...). \end{aligned} \tag{2.12}$$

Since we used DLSR camera and variable focal-length lens in image acquisition system, the pre-calibration before each image acquisition becomes very difficult and impractical. Therefore, in the following chapters, an estimation method of camera parameters and poses will be developed, that takes radial distortion into account as well.

## 2.3 Summary

In this chapter, the background of existing 2-D/3-D imaging methods for plant phenotyping has been presented, including 2-D digitalizing and visualization, 3-D sensor techniques, as well as the advantages and limitations of these techniques. Although the development of advanced 3-D imaging techniques makes 3-D spatial data available for plant structure analysis, these 3-D imaging methods do not achieve adequate accuracy, reasonable performance or a more affordable system cost, and 3-D phenotypic features extraction are still limited to those simple ones. Therefore, more efforts still need to be made to achieve a cost-effective 3-D imaging resolution that can produce accurate, complete 3-D spatial data and can extract in-depth 3-D structural features with the accuracy, processing speed, and reliability desired by plant biologists.

# Chapter 3

# Multi-view Image Acquisition Methodolody

*"Garbage in, garbage out"* – this saying emphasises the importance of quality of raw data acquisition. Although commencing as a relatively simple structure, the plant rapidly grows and becomes more complex leading to an overlapping self-occluding 3-D structure. A DSLR camera with variable focal-length lens has the ability to capture high-resolution 2-D image under variable environmental conditions. In this chapter, a multi-view image acquisition system is proposed to capture high quality plant images only using low-cost photography equipment, without pre-calibration. Additionally, typical plant phenotyping experiments and evaluation criterion are introduced.

## 3.1 Plants Growth

In my research, all plants (Arabidopsis, grass, wheat, maize, clover, etc.) were grown in appropriately sized pots (depending on species and developmental stage), with adequate watering (approximately 70% field capacity) at the Botany Gardens Glasshouse or National Plant Phenotyping Center (NPPC) of IBERS of Aberystwyth University under a semi-controlled growing environment (Greenhouse) where temperature was set between 15°C and 20°C with supplementary lighting. Plants were moved temporarily on to a turn table during imaging.

## 3.2 Multi-view 2-D Image Acquisition

My image acquisition scenario was inspired by the results reported from a well-known MVS (Multi-View Stereo) benchmark and evaluation project (Seitz *et al.*, 2006a), where the multi-view images were captured using the Stanford Spherical Gantry (Levoy, 2002) – a robotic arm that can be positioned on a one-meter radius sphere to an accuracy of approximately 0.01 degrees and enables a camera to be moved on a sphere to specified latitude/longitude angles (see Fig. 3.1). The gantry cost $130,000 to design and build(Levoy, 2002).

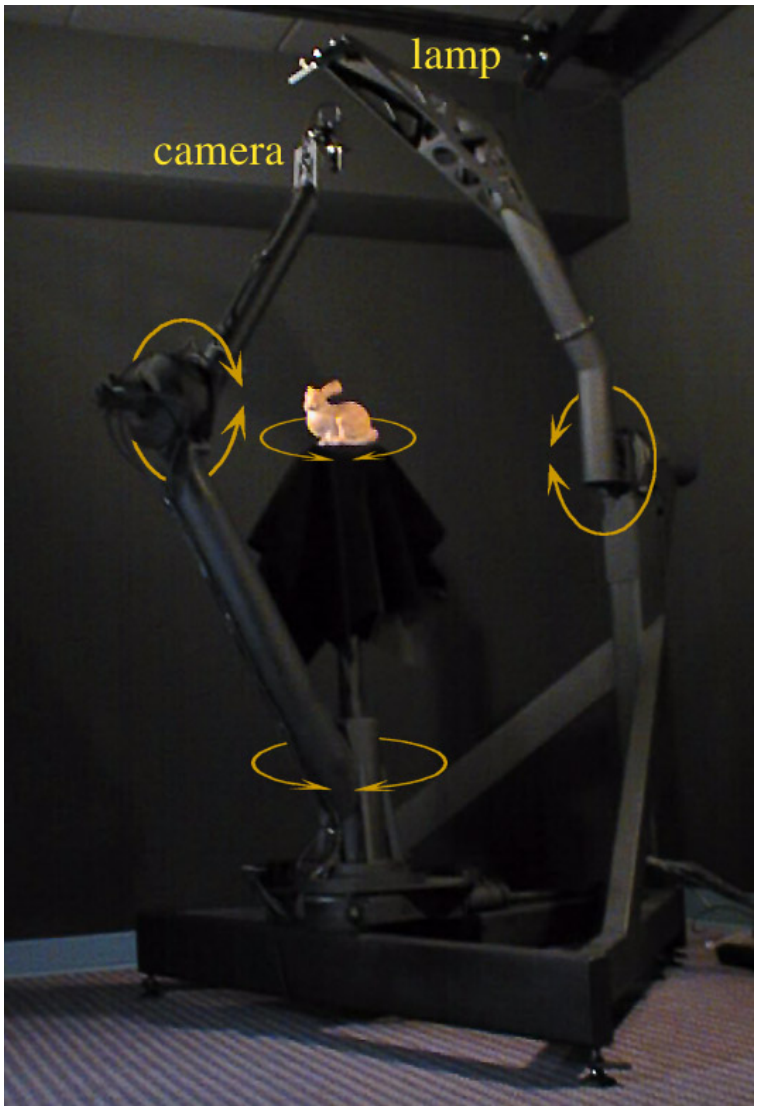



Figure 3.1: Stanford spherical gantry. Image courtesy of (Levoy, 2002)

In (Seitz *et al.*, 2006a), the target object sat on a stationary platform near the center of the gantry sphere and was lit by three spotlights, and multi-view images were captured using a CCD camera with a resolution of 640×480 pixels attached to the tip of the gantry arm. At this resolution, a pixel in the image spans roughly 0.25mm on the surface of the object (the temple object is 10cm×16cm×8cm, and the dino is 7cm×9cm×7cm, see Fig. 3.2). The camera was calibrated by imaging a planar calibration grid from 68 viewpoints over the hemisphere and using calibration toolbox (Bouget, 2012) to compute intrinsic and extrinsic parameters. From these parameters, the authors computed the camera's translational and rotational offset relative to the tip of the gantry arm, which determine the camera's position and orientation as a function of any desired arm position.

Figure 3.2: the temple and dino. The temple object is 10cm×16cm×8cm, and the dino is 7cm×9cm×7cm. Image courtesy of (Seitz *et al.*, 2006a).

From the resulting images, the authors created three datasets for each object, corresponding to a full hemisphere, a single ring around the object, and a sparsely sampled ring (see Fig. 3.3). The results of a quantitative comparison of state-of-the-art multi-view stereo reconstruction algorithms on six benchmark datasets (Seitz *et al.*, 2006a) are as follows:

(1) Most methods consistently achieve sub-millimeter accuracy with very few outliers even from the images captured at a low resolution of 640×480 pixels;

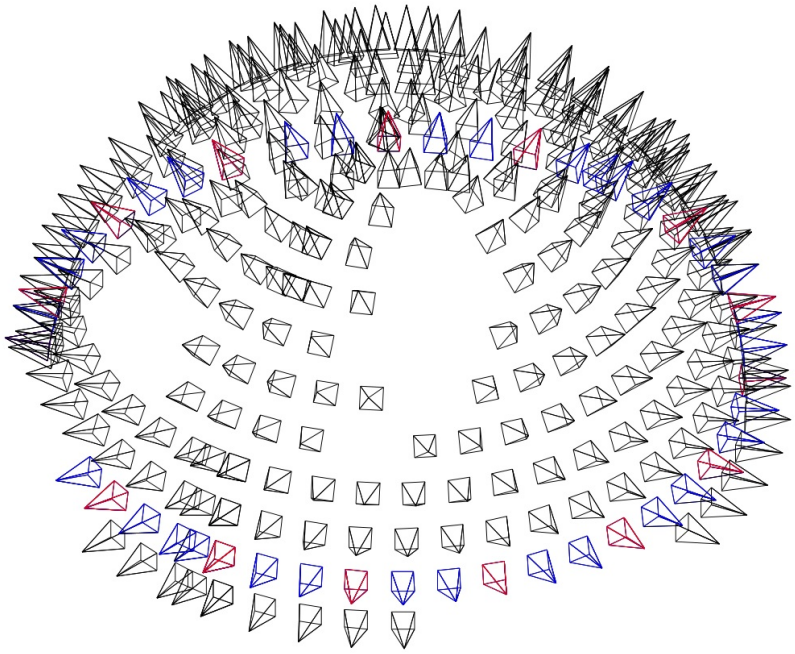


Figure 3.3: the 317 camera positions and orientations for the temple dataset. The gaps are due to shadows. The 47 cameras corresponding to the ring dataset are shown in blue and red, and the 16 sparse ring cameras only in red. Image courtesy of (Seitz *et al.*, 2006a).

(2) Accuracy in all methods decreased slightly with fewer images on the temple datasets, e.g. for PMVS (Furukawa & Ponce, 2010) on three different sampling datasets of the temple, where accuracy is 0.49mm, 0.47mm and 0.63mm from the full dataset (317 images) , ring dataset (47 images) and sparse ring dataset (16 images) respectively;

(3) Accuracy on the dino datasets surprisingly show the opposite trend, with most methods doing better on the ring than on the full dino dataset. Due to the fact that the dino lacks textures on its surface, regularization likely plays a more central role.

Therefore, considering that plants are non-volumetric objects with complex surfaces and different shape sizes (ranging from a few centimeters to one meter), I designed an image acquisition method (shown in Fig. 3.5) that only used off-the-shelf consumer-grade equipment − a common high-resolution digital camera with zoom lens and a turn-table which enables the plants to rotate by 360°−

to capture 1∼2 continuous ring dense image sequences (usually 1 ring image sequence is enough for most small plants). I have captured tens of thousands of plant images without pre-calibration required in advance, and built up a dataset of 3-D plant point clouds of various plants including *Arabidopsis thaliana*, oat, wheat, maize, forage grass (*Lolium*), clover, oilseed rape, physalis and brassica.

Fig. 3.4 and Fig. 3.5 illustrate the principle of multi-view 2-D image acquisition of plants. Fig. 3.4 shows the setup with a digital camera, a turn-table, two lighters and a computer. Fig. 3.5 shows the schematic lay-out of the setup. To reduce the shadow (usually caused by self-occlusion of plants) and obtain even illumination and uniform background (reduce the noise), two diffuse lighting rigs and a black backdrop were used.

The camera can acquire images in landscape or portrait format, and some sample images of observed plants are shown in Fig. 3.6 and Fig. 3.7, respectively.


Figure 3.4: multi-view 2-D imaging setup.

The procedure of image acquisition is summarised in the following steps:

(1) Place the plant at the centre of the turn-table, adjust the camera focus length (or the distance between plant and tripod) to ensure the whole plant is located at the centre and fills most of the frame, but with margins on all sides.

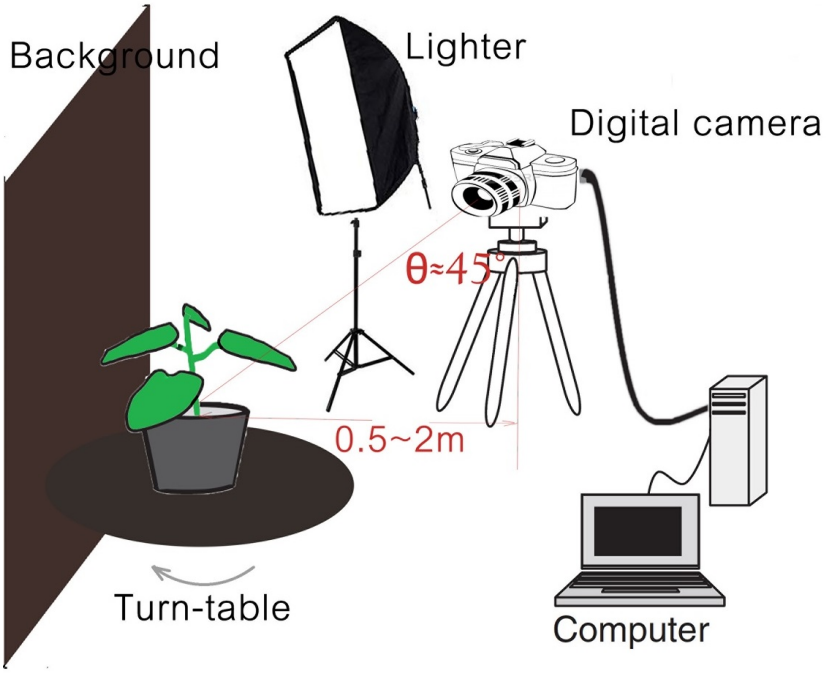



Figure 3.5: schematic lay-out of multi-view 2-D imaging setup.

(2) The camera should point at the lower section of the main stem of the plant rather than the top section or leaf surface, and then start auto-focusing.

(3) Switch on turn-table to start rotation. Once one hemisphere has passed and the plant has stopped vibrating, trigger the camera to take images under "continuous shooting mode".

(4) Stop shooting when the plant has rotated over a further 360°.

(5) Transfer the captured image sequence (about 100∼200 pictures) to the computer.

In order to obtain the best photographic quality (clarity and accuracy) record of various plants with different size/form, some parameters (e.g. aperture, depth of field, ISO, shutter speed) of the camera were adjusted to appropriate values (see below), using the autofocus (AF) but disabling the flash. The camera's configuration and shutter release are either adjusted manually or controlled by a commercial remote-trigger and configuration software, "DSLR Remote Pro"(Systems, 2014), via a camera


Figure 3.6: sample images of observed plants with high-resolution of 3456×2304 pixels captured by Canon DSLR camera (600D). Brassica (left) and *Arabidopsis* (right). The scale bar (white line) indicates 5cm in images.

USB cable. The captured images are transferred from the camera to the computer via the USB cable.

In my system, plant images were captured using a Canon DSLR camera (600D) with 18∼55mm lens, or Nikon DSLR camera (D700) with 28∼105mm lens. The black turn-table used in my system was a turn-table (costs ∼ $20) for 360° rotating display, whose diameter is 25.4 cm. The rotation speed was approximately one circle per 40 seconds (the speed basically determines the number of acquisition images per circle). For camera configuration in most image acquisitions, *ISO* was fixed at "200" to adapt the lights, and *Shutter speed* was selected to be "1/12" to match the rotation speed of the turn-table, *Aperture* was adjusted at "f/13" (if plant are young, or with small/medium shape, e.g. *Arabidopsis*, grass), or "f/16" (if the plants are with big shape, e.g. wheat, maize) for obtain the best Depth of Field (DoF).

The captured images were stored in JPEG format with a high-resolution of 3456×2304 (Canon 600D) or 3696×2448 (Nikon D700) pixels and 24-bit

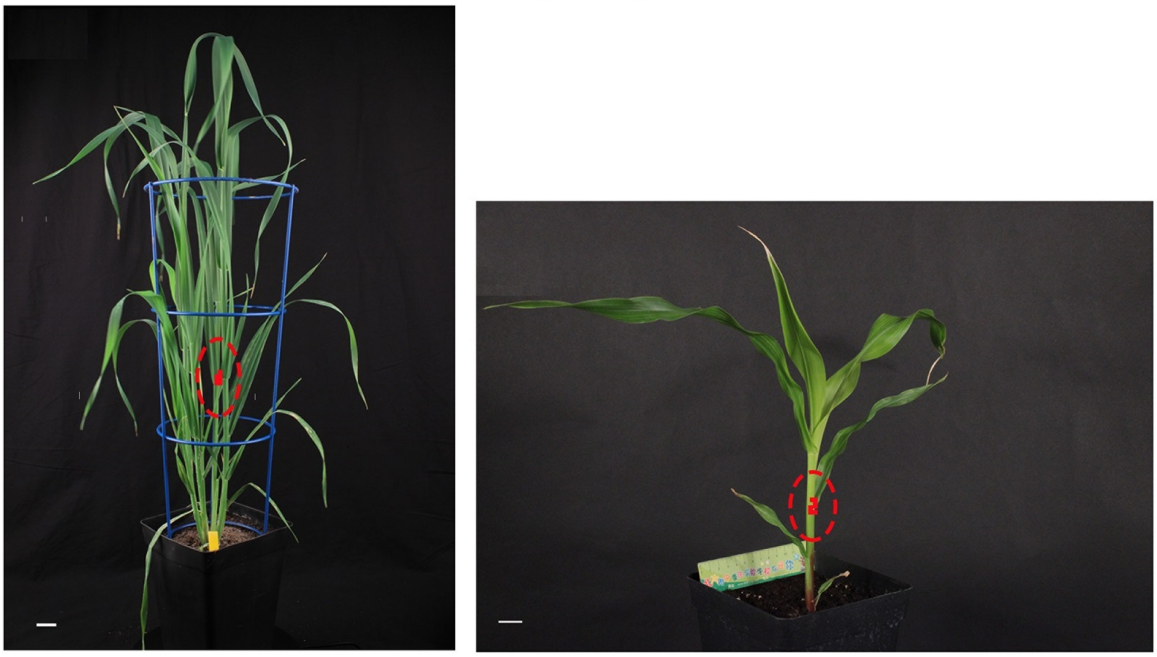


Figure 3.7: sample images of observed plants with high-resolution of 3456×2304 pixels captured by Canon DSLR camera (600D). wheat (left) and maize (right). The scale bar (white line) indicates 5cm in images.

color depth. Although JPEG is a lossy image compression format, it doesn't degrade the quality of experimental result, therefore, I used JPEG to store the images instead of RAW or PNG. Videos were also captured at a resolution of 1920×1024 pixels in AVI format in some experiment.

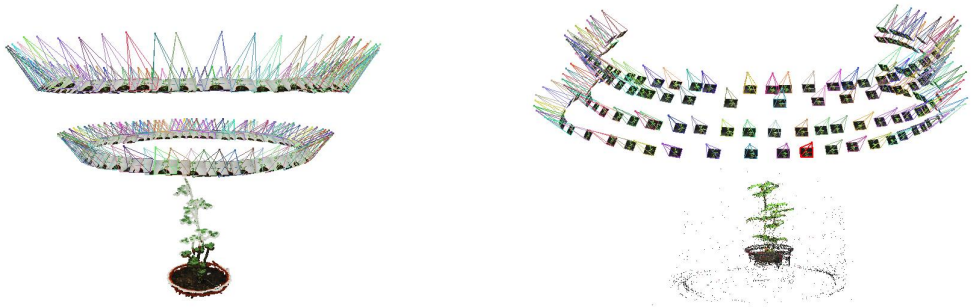
The computer used for data capture and storage was a Dell laptop, Precision M6700 with Intel i7 2.7GHz CPU, 16GB RAM and NVIDIA GPU graphics card, running 64 bit Windows 7.

Although the acquisition was manual, it used multi-view imaging to resolve the occlusion problem as much as possible; this is in contrast to a commercial high-throughput phenotyping platform − e.g. LemnaTec − where automated plant handling and imaging have been implemented. The LemnaTec system captures 2-D RGB images from only a limited number of directions (top-view, user-defined side-views between 0° and 90°) using digital cameras installed at fixed positions with pre-calibrated parameters; it cannot handle the occlusion problem well, but only different shape/size of various plants.

The proposed method can be adjusted in the sense of the camera distance,

pose, focus, lens, depth of field and exposure time to obtain high quality and clear images with an optimal resolution for a diverse set of plants. This means that only basic photography skills are required for image capturing, and no extra camera pre-calibration is needed. The average time of data acquisition of a plant was in the range of 1∼2 minutes depending on its complexity.

For the plant with bigger shape or severe self-occlusions, I used 2∼3 cameras to simultaneously shoot the plant with 2∼3 different heights and poses so as to improve the quality of images. The illustration is shown in Fig. 3.8.



(a) capturing a plant with two cameras

(b) capturing a plant with three cameras

Figure 3.8: illustration of image capturing using multiple cameras with different heights and poses.

## 3.3 Plant Phenotyping Experiment

As mentioned in the previous chapters on typical plant phenotyping experiments, investigated plants are usually grown in semi-controlled conditions (e.g. glasshouses) and subjected to different environmental conditions and stresses (e.g. drought, salt, temperature, etc.) with the primary aim of monitoring and collecting their phenotypic response using various measurements (Granier & Tardieu, 2009)(Schurr *et al.*, 2000). Common plant morphological traits of interest include parameters, such as main stem height, size and inclination, petiole length and initiation angle, and leaf width, length, inclination, thickness, area, and biomass, etc. (Furbank & Tester, 2011)(Granier & Tardieu, 2009)(Schurr *et al.*, 2000)(Vos *et al.*, 2010)

Image analysis tools should be capable of quantifying and describing such phenotypic traits. Therefore, in this project, a typical experiment of plant

phenotyping usually need to capture a group of plants, which could have same strain (also known as "ecotype") but grown in different conditions to subject to different environmental conditions and stresses (e.g. drought, salt, temperature, etc.), or could have different ecotypes grown in the same conditions.

Fig. 3.9 shows sample *Arabidopsis* images in a phenotyping experiment. There were more than 100 different ecotypes of *Arabidopsis* grown in the same controlled conditions, and were captured by the proposed multi-view image acquisition system every 1∼2 days over a few weeks. In another phenotyping experiment, $8 \times 3$ grasses (8 different ecotype, 3 sample per ecotype) were captured every day (biologists hope to analyse different rates of grass growth), see Fig. 3.10.

## 3.4 Ground Truth and Evaluation Criterion

Due to limitations of existing 3-D techniques, it is difficult to obtain the ground truth of full 3-D architecture of an observed plant. Therefore, some phenotypic parameters can be only manually measured as ground truth by biologists, and such measurements are usually destructive and take place at harvest stage.

Linear regression or correlation can reveal whether one measurement variable is associated with another measurement variable. For biologists, the results of statistical analysis (for example, correlation and linear regression) are often used to evaluate the outcome of measurement experiments, e.g., whether or not some variable has an effect, whether variables are related, whether differences among groups of observations are the same or different.

A sample of evaluation on quality of 3-D measurements is illustrated in Fig. 5.11 (see Chapter 5), where the high correlation of $R^2 = 0.97$ (close to 1) between the ground truth and measured results indicates that 3-D measurement results are reliable or meaningful.

## 3.5 Summary

In this chapter, a cost-effective multi-view 2-D image acquisition system is proposed to resolve the plant's structure self-occlusion problem as much as possible,

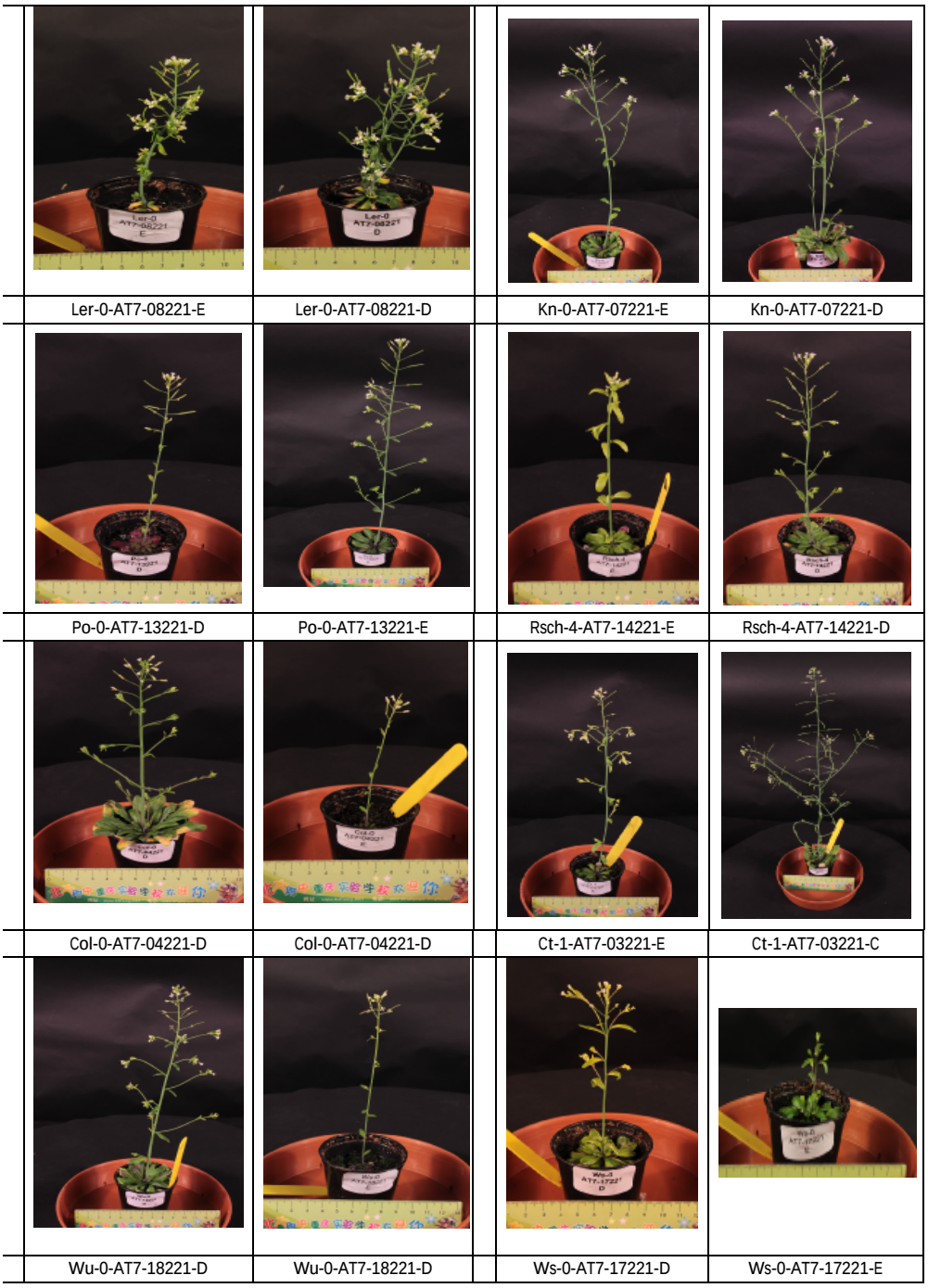



Figure 3.9: Sample images of *Arabidopsis* captured at 27/03/2014 in the NPPC. There are more than 50 different ecotype *Arabidopsis* grown in the same controlled conditions in a phenotyping experiment, which were captured every 1∼2 days through their whole growth stages.

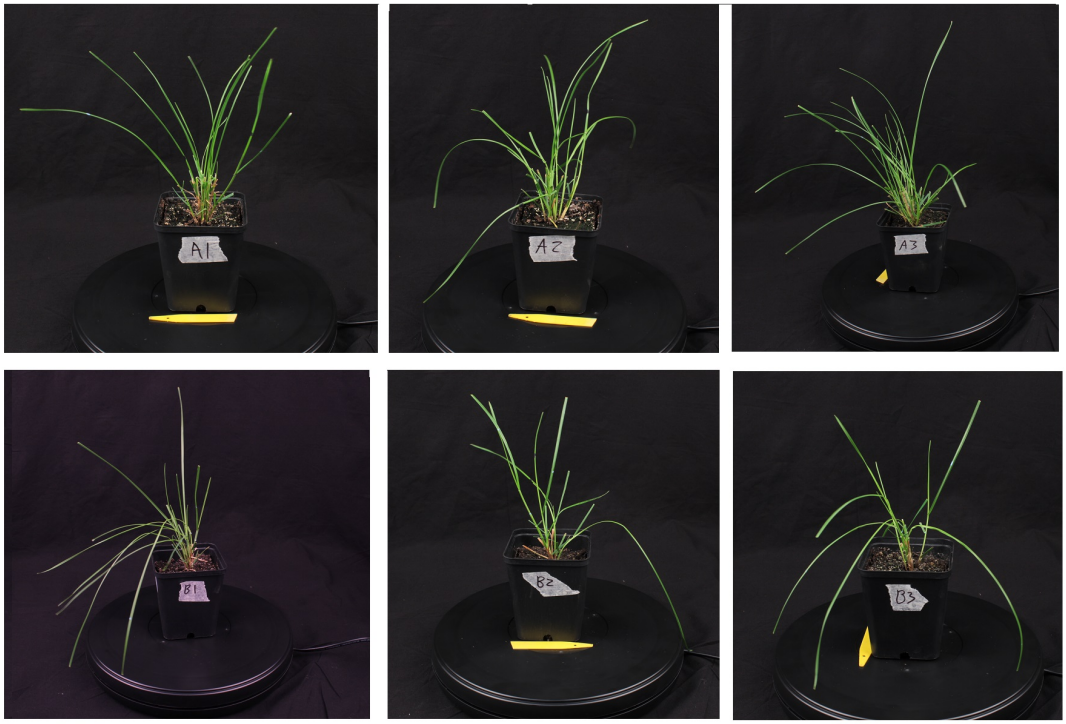



Figure 3.10: 6 sample images of grass captured at 2/07/2014 in the greenhouse of the NPPC with the aim to analyse the rate of the grass growth. There were total 24 grasses (8 different ecotype) in the phenotyping experiment.

which consists of some common digital photograph equipments (such as DLSR camera, tripod, lighting, uniform background curtain and a computer), and is capable of capturing high-resolution images of variety plants under their different growth stages. The main advantages of multi-view 2-D image acquisition system are its low-cost, flexibility, as well as no pre-calibration required.

# Chapter 4

# Camera Parameters and Pose Estimation from Multi-view Images

In the multi-view 2-D image acquisition system, the subject (plant and its pot) is placed approximately at the center of a turntable, and the camera positioned and adjusted to capture images of the subject. Thus, the camera can be considered to be "moving" relative to the subject. However, the motion of camera cannot be simply considered as pure circular motion because the plant is an irregular object, and it is difficult to accurately place the plant at the center of the turntable. In this chapter, a general efficient method has been developed to estimate the camera motion (poses and locations) from uncalibrated multi-view plant images. The method is based on Structure from Motion techniques, and experimental results have shown the validity of the proposed method.

## 4.1 Introduction

Structure from Motion (SFM) (Ullman, 1979) has been an active research area in computer vision for decades, and numerous applications, e.g. 3-D reconstruction, stereo vision, augmented reality, camera motion tracking, and autonomous mobile robot localization/navigation, are benefiting from this research due to the simplicity in concept and universal applicability.

The SFM technique refers to the process of simultaneously estimating the 3-D geometry of a scene (*structure*) and the cameras' poses (*motion*) from a sequence of images taken with a moving camera, and often integrates a variety of methods and tools from multi-view geometry, image processing and statistical regression. SFM can be defined as the problem (see Fig. 4.1) (Hartley & Zisserman, 2004): given $m$ images of $n$ fixed 3-D points, $x_{ij} = \mathbf{P}_i X_j, (i = 1, 2..., m; j = 1, 2..., n)$, how to estimate $m$ projection matrices $P_i$ and $n$ 3-D points $X_j$ from the $mn$ correspondences $x_{ij}$ ?

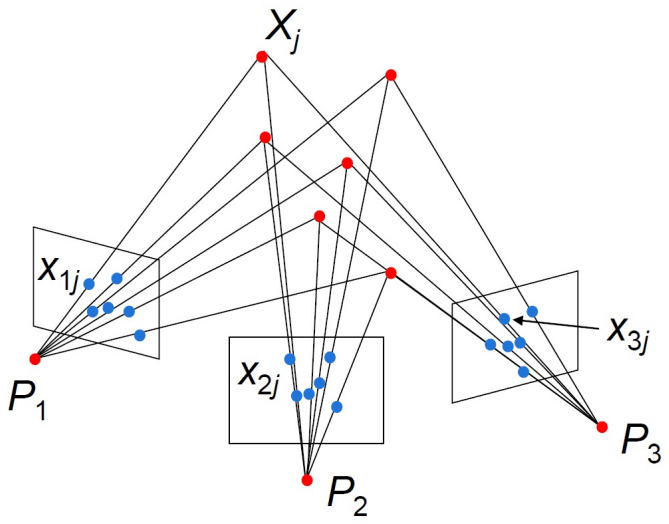



Figure 4.1: SFM is to estimate the cameras' parameters/poses and 3-D points from multiple images.

A typical pipeline of incremental SFM application consists of several processes, as follows:

(1) Feature points are first detected in each image of input photographs. Fig. 4.2 shows a real example (Snavely *et al.*, 2006), whereas scale-invariant feature transform (SIFT) (Lowe, 2004) detector was applied to detect feature points from an unordered set of city scene images.

(2) Pairwise feature points are matched using SIFT descriptors. Matched feature points between each pair of images often include incorrect correspondences (outliers) due to similar or repeated structures of the scenes. See the example shown in Fig. 4.3(a).

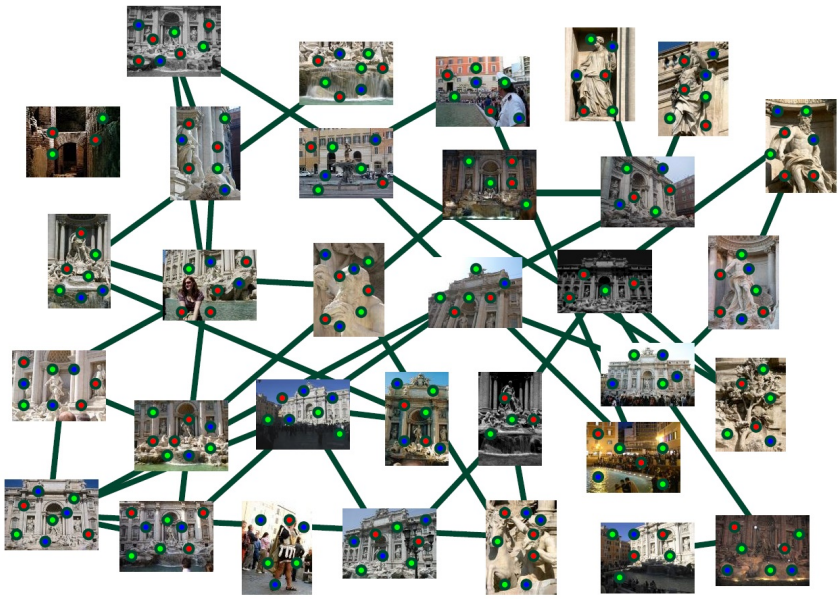
(3) Matching is refined using Random Sample Consensus (RANSAC) (Fischler & Bolles, 1981) and two-view geometry to estimate the fundamental/essential matrices between the pair of images. The outliers can be removed as shown in Fig. 4.3(b).

(4) A 3-D model is reconstructed by initializing from the good two-view geometry, repeatedly adding the next matched image, triangulating the corresponding features, and refining the structure and motion by minimizing re-projection errors. This incremental processing is illustrated as Fig. 4.4.

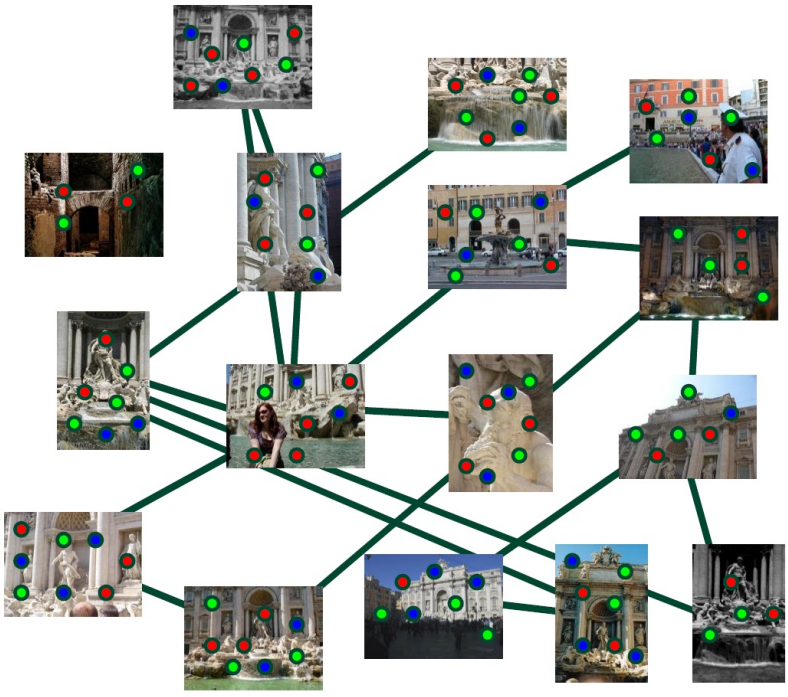
(5) Refining the final structure and motion by minimizing the sum of re-projection errors.

Figure 4.2: Input an unordered set of images, detecting SIFT features of each image. Image courtesy of (Snavely *et al.*, 2006).

In the pipeline of SFM, bundle adjustment is an effective non-linear optimization processing to simultaneously refine the cameras' parameters and 3-D structures. The camera projection matrix is denoted by $\mathbf{P} = \mathbf{K}[\mathbf{R}|t]$ , where $\mathbf{K}$ is the intrinsic camera calibration matrix, $\mathbf{R}$ is the rotation matrix, and $t$ is the translation vector of a camera (Hartley & Zisserman, 2004). Where $m$ is the number of cameras and $n$ is the number of 3-D points, bundle adjustment



(a) Pairwise matching SIFT features.



(b) Refine matching using RANSAC and two-view geometry.

Figure 4.3: (a) SIFT features matching between each pair of images. (b) Refine matching using RANSAC and two-view geometry. Image courtesy of (Snavely *et al.*, 2006).

Figure 4.4: Construct and refine the 3-D structures and cameras' poses incrementally. Image courtesy of (Snavely *et al.*, 2006).

(Triggs *et al.*, 1999) aims to minimize the sum of the re-projection errors.

$$min \sum_{i=1}^{n} \sum_{j=1}^{m} D(x_{ij}, \mathbf{P}_i X_j)^2, \tag{4.1}$$

where $x_{ij}$ is the projection of 3-D point $X_j$ onto the $i$-th image, $\mathbf{P}_i$ is the projection matrix of the $i$-th camera, see Fig. 4.5.

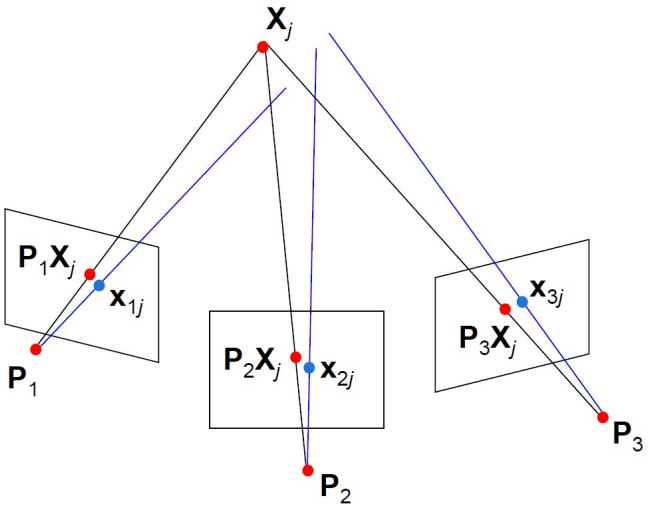



Figure 4.5: Bundle adjustment is a non-linear optimization method for refining structure and motion by minimizing the sum of re-projection errors.

Although the SFM pipeline is robust and able to perform well in most cases, the time complexity of such an incremental SFM algorithm is commonly known to be $O(n^4)$ for $n$ images (Wu, 2013), and this high computation cost impedes the application of incremental SFM on large scale image datasets, where the

3-D reconstruction usually needs to run a few hours or tens of hours even with the high performance computers (Shan *et al.*, 2014; Wu, 2013; Agarwal *et al.*, 2011; Agarwal *et al.*, 2010; Snavely *et al.*, 2010; Wu *et al.*, 2011).

## 4.2 The Proposed Method

As mentioned in previous chapters, a sequence of images contains about 60∼120 images with high resolution (e.g. 3456×2304 pixels), and a typical experiment of plant phenotyping often needs to capture thousands or tens of thousands of images over a few weeks. Therefore, there is an existing great challenge in computing abilities for most biology labs.

In order to estimate the cameras' parameters and poses from multi-view plant images within a reasonable time frame, an efficient incremental SFM approach has been developed to address the challenge. The framework of the proposed method is illustrated as Fig. 4.6 and described in the following sections.

### 4.2.1 Image Features Detection and Matching

The first stage of the framework is to build a set of corresponding points between each pair of images by local invariant features detection and matching.

One of the most popular local invariant feature is SIFT, as presented by (Lowe, 2004), which detects key points by searching for maximum in scale-space, constructed with derivative of Gaussian filters, and then a unique descriptor is constructed to represent each key points based on a histogram of gradient directions of points nearby. SIFT features are invariant to image scale and rotation, and partially invariant to intensity and contrast changes (Lowe, 2004). However, the main drawback of SIFT is that the computational complexity of the algorithm increases rapidly with the number of key points, especially at the matching step due to the high dimensionality of SIFT feature descriptor.

Alternative features can be used as well, so long as they can be matched uniquely between two images, such as Harris (Stephens & Mike, 1988), SURF (Bay *et al.*, 2008), ASIFT (Affine-SIFT) (Morel & Yu, 2009), DAISY (Tola

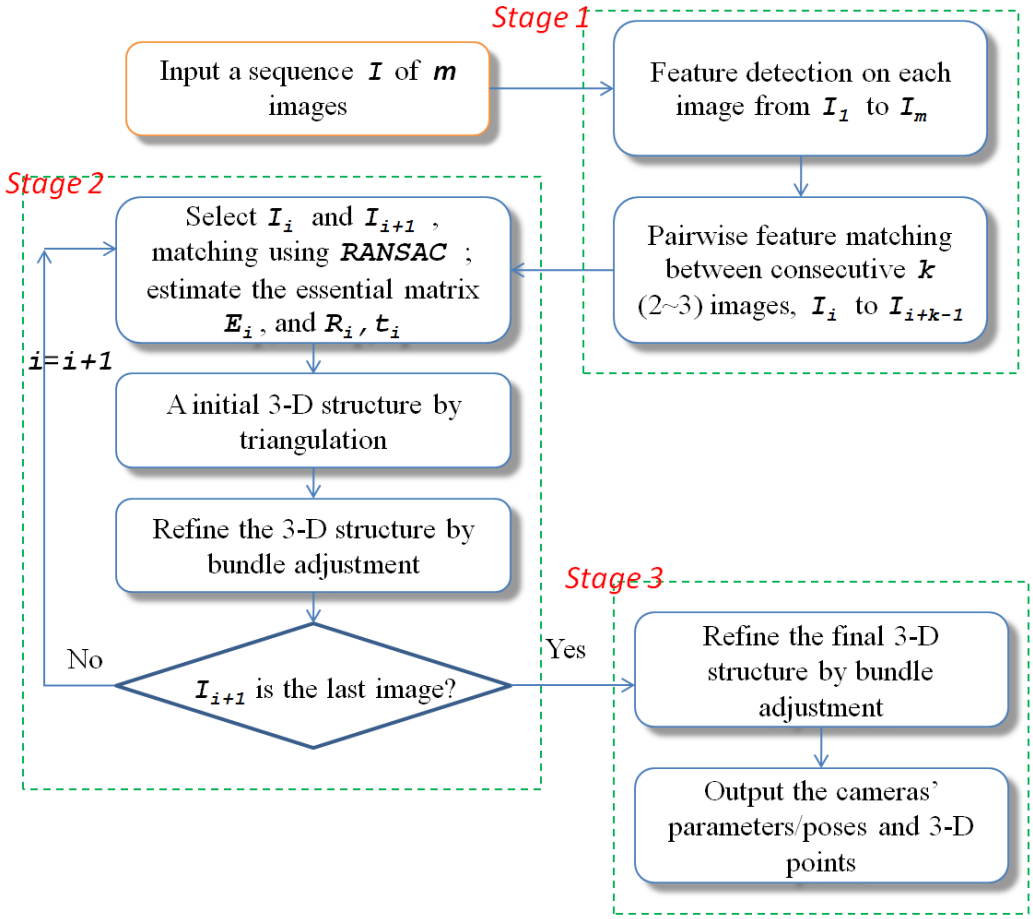



Figure 4.6: The workflow of the proposed method.

*et al.*, 2010) and AKAZE (Alcantarilla *et al.*, 2013). For example, the ASIFT detector/descriptor is used as an alternative in some experiments, since the authors in (Morel & Yu, 2009) argued that ASIFT was better than other popular feature algorithms as SIFT, MSER and Harris-Affine on various datasets. Although many more features can indeed be detected and matched, ASIFT did not improve the experimental results. Therefore, SIFT is mainly tested in this project due to its repeatability and accuracy.

Generally, an image with higher resolution contains many more features. A high-resolution image of a plant often contains hundreds or thousands of SIFT features. In order to overcome the computational complexity of SIFT, SiftGPU (Wu, 2011) is applied in my method to speed up SIFT features detection by GPU hardware (see (Wu, 2011) for details).

For the same reason, in order to speed up the SIFT feature matching, the library of Approximate Nearest Neighbor (ANN) (Arya *et al.*, 1998) is used to improve the speed of feature searching and matching (refer to (Arya *et al.*, 1998) for details).

In this stage, those matched features could contain outliers − two matched features with incorrect geometry correspondences.

### 4.2.2 Estimation of Camera Parameter and Pose

The second stage of SFM is to estimate camera parameters and pose by use of related techniques.

Given a short baseline sequence of images $I$ with $m$ images, $I = \{I_i | i = 1, 2, ..., m\}$, where $I_i$ represents the $i$-th image (stored as JPEG file). The $i$-th camera's parameters $I_i$ is denoted as $C_i = \{K_i, R_i, t_i\}$, where $K_i$ is the intrinsic matrix, $R_i$ is the rotation matrix, and $t_i$ is the translation vector of the $i$-th camera.

First, the value of the focal length embedded into the EXIF tags of $I_i$ can be extracted to compute $K_i$. For example, in most experiments of this project, the plant image files were captured in JPEG format with a high-resolution of 3456×2304 pixels (width×height) using a Canon 600D camera equipped with a variable 18∼55$mm$ focal-length lens, and I can deduce the CCD size of Canon 600D from digital camera databases − it is 22.3mm wide×14.9mm height. If I extract the focal length $F_{mm}$ from EXIF tag in sequence image $I_i$ of a plant, supposed the $F_{mm}$ is 35$mm$, then based on the equations (2.8) and (2.9), the intrinsic matrix $K_i$ could be computed as:

$$K_i = \begin{bmatrix} 5424 & 0 & 1728 \\ 0 & 5424 & 1152 \\ 0 & 0 & 1 \end{bmatrix}.$$

This intrinsic matrix $K_i$ is the same for all images of the sequence, since the focal length of the camera was fixed during the period of one data acquisition.

Second, a pair of images $I_i, I_{i+1}$ (two neighboring images, which are supposed to have a maximal number of matching points) are chosen, then outliers

of matching points between $I_i$ and $I_{i+1}$ are removed by RANSAC – during each RANSAC iteration, a candidate of essential matrix $E_i$ between $I_i$ and $I_{i+1}$ is estimated as the consistent constraint using five-points algorithm (Nistér, 2004) – once the essential matrix $E_i$ is estimated, the rotation matrix $R_i$ and translation vector $t_i$ (relative pose of the two cameras) can be extracted from $E_i$, and then an initial set of 3-D points is produced by triangulating(Hartley & Zisserman, 2004). The *structure* and *motion* of the pair of images $I_i, I_{i+1}$ – the initial set of 3-D points and the essential matrix – are further refined by bundle adjustment (Triggs *et al.*, 1999). After that, a new neighboring image $I_{i+2}$ is added into the current 3-D structure, repeatedly running above processes (RANSAC, triangulating and bundle adjustment) until all images are processed. See Fig. 4.7.

Finally, a global bundle adjustment is applied to refine the entire model – 3-D points and cameras' poses. The library of SBA (Lourakis & Argyros, 2009), a software package for generic sparse bundle adjustment, is applied in the bundle adjustment procedure.

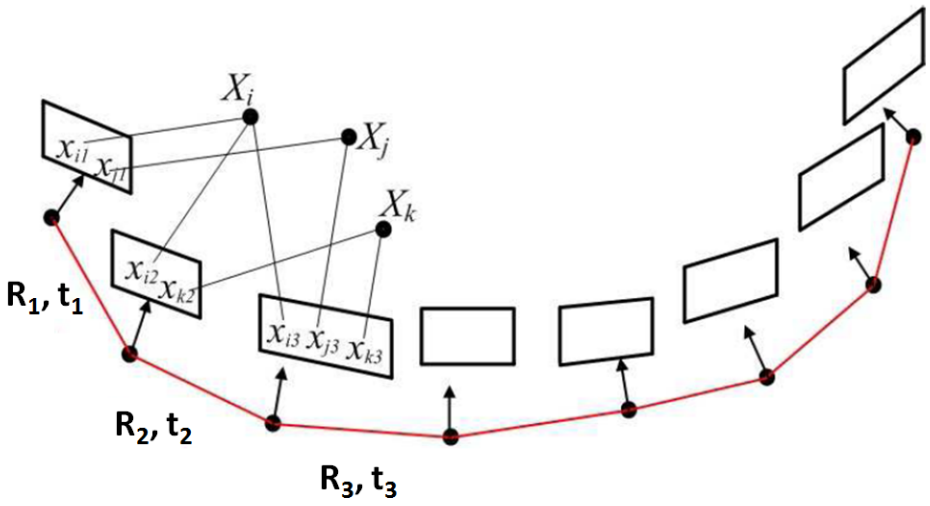



Figure 4.7: Construct and refine the 3-D structures and cameras' poses incrementally.

The proposed method is inspired by the early work of Snavely *et al.*. (Snavely *et al.*, 2008), an open source software *bundler*. Compared with other existing

SFM systems, such as *bundler*(Snavely *et al.*, 2008), *openMVG*(Moulon, 2012), the performance of the proposed SFM approach has some improvement as follows:

(1) To avoid pairwise feature matching in $n$ images, I sort the images as an ordered sequence according to incremental image timestamps so that SIFT matching is incrementally performed only between two neighboring images, thereby dramatically reducing the computational cost from $O(n^2)$ to $O(n)$ in the number n of the images for key points matching;

(2) A SiftGPU (Wu, 2011) algorithm is applied to speed up the SIFT detection on the computer with GPU display card, which is approximately 10 times faster than CPU;

(3) For computers without GPU hardware, as alternative means of speeding up implementation, other efficient feature detection/matching algorithms, such as SURF (Bay *et al.*, 2008), DAISY (Tola *et al.*, 2010) and AKAZE (Alcantarilla *et al.*, 2013), can be used in the step of feature detection and matching.

## 4.3 Experimental Results and Discussion

To evaluate the accuracy of the proposed method, the experiments were firstly conducted on the well-known MVS datasets − Middlebury dino and temple − which were created in (Seitz *et al.*, 2006a), where the multi-view images were acquired using the Stanford Spherical Gantry, meanwhile, the camera's parameters and poses were computed by the authors with a planar calibration grid and the camera's translational and rotational offset relative to the tip of gantry arm (see Chapter 3). Secondly, the experiments were carried on the plant images using the data acquisition system proposed in Chapter 2, where multi-view plant images were captured using a low-cost turn-table and a Canon digital camera (Canon 600D) with 18∼55$mm$ focal lens, and were stored in JPEG format with high-resolution 3184 × 2120 pixels.

All experiments were run on an Intel i7 laptop, Dell Precision M6700 with 16G RAM and Nvidia graphics card.

### 4.3.1 Evaluation of the Accuracy of the Proposed Method

The experiments were conducted on the sparse ring datasets of Middlebury dino and temple (Seitz *et al.*, 2006b), see Fig. 3.3.

Because of the different reference coordinate between Middlebury dataset and the proposed method, for simplicity but without loss of generality, a sequence of 18 consecutive images and a sequence of 22 consecutive images were respectively selected from Middlebury temple and dino to estimate the camera's poses with the proposed SFM method, see Fig. 4.8 and Fig. 4.9.

In the experiments, camera's poses were estimated from two sequences and the rotation matrices between two neighboring images were then computed. The cameras' poses are illustrated in Fig. 4.10 and Fig. 4.11 respectively. The angles of step of the camera rotation were compared with the ground truth so as to evaluate the accuracy of the proposed SFM method.

The experimental results in Table. 4.1 and Table. 4.2 show the ground truth of step of the camera rotation is 7.66° and 7.83° , and the root mean square error (RMSE) of the estimated angle of rotation step is 0.26° and 0.23° respectively − It shows that the accuracy of the proposed SFM method is high.

### 4.3.2 Key Point Detection and Matching on Plant Images

This experiment was designed to verify the performance of key point detection and matching in the proposed method. This experiment was conducted on a sequence of 54 images of a Brassica. SIFT features were used to detect and match $m$ consecutive images, $m = 5$ was selected in the experiment. Five consecutive images (48th, 49th, 50th, 51st and 52nd) were selected randomly as illustration examples, see Fig. 4.14(a).

The experimental results show that the 48th, 49th, 50th, 51st, 52nd images had *3905, 3819, 3723, 3632, 3565* SIFT features, respectively; the number of matching SIFT features (including outliers) for the pairs, 48th-49th, 48th-50th, 48th-51st and 48th-52nd, is *1029, 359, 148, 88*, respectively, see Fig. 5.3(a) and Fig. 4.14(c).

After outliers are removed by RANSAC processing, the number of inliers is

Figure 4.8: a sequence of 18 consecutive images selected from the dataset of Middlebury temple.

*912, 290, 97, 0*, respectively, see Fig. 4.13(b) and Fig. 4.13(c). These results indicate that the number of available matching SIFT feature will significantly

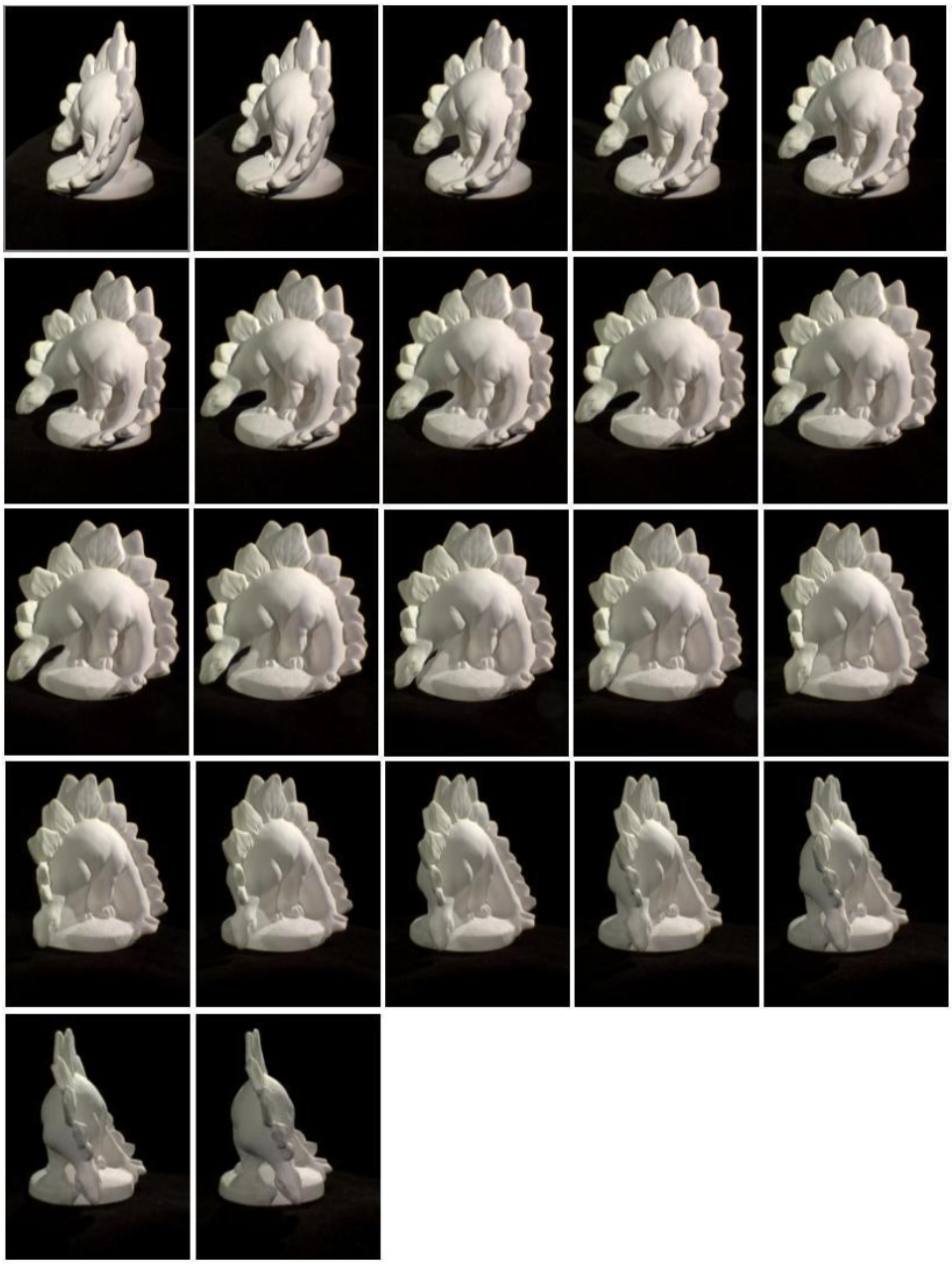


Figure 4.9: a sequence of 22 consecutive images selected from the dataset of Middlebury dino.

decrease when the degree of rotation increases, and, therefore, sequence matching between at most 3∼4 consecutive images is reasonable and effective.

Fig. 4.12 and Fig. 4.13 indicate that leaves often contain outliers, and much more SIFT features were detected and matched (including inliers matching) on the soil or turntable rather than the leaves. These results suggest that

Figure 4.10: estimation of cameras' poses from the dataset of Middlebury temple.

Figure 4.11: estimation of cameras' poses from the dataset of Middlebury dino.

the appearances of the pot and turntable are very helpful for SIFT features detection and matching, especially when the plant leaves lack surface textures.

Fig. 4.14 illustrates the epipolar geometry (Hartley & Zisserman, 2004) between consecutive images, which can be computed easily from the inliers obtained from the feature matching process.

| Item | Ground truth | Estimated value | Relative error(%) |
|---|---|---|---|
| angle1 of rotation step | 7.66 | 7.71 | 0.72% |
| angle2 of rotation step | 7.66 | 7.66 | 0.00% |
| angle3 of rotation step | 7.66 | 7.58 | -0.99% |
| angle4 of rotation step | 7.66 | 7.69 | 0.34% |
| angle5 of rotation step | 7.66 | 7.51 | -1.88% |
| angle6 of rotation step | 7.66 | 7.31 | -4.54% |
| angle7 of rotation step | 7.66 | 7.22 | -5.77% |
| angle8 of rotation step | 7.66 | 7.21 | -5.93% |
| angle9 of rotation step | 7.66 | 7.75 | 1.20% |
| angle10 of rotation step | 7.66 | 8.13 | 6.13% |
| angle11 of rotation step | 7.66 | 7.87 | 2.82% |
| angle12 of rotation step | 7.66 | 7.79 | 1.67% |
| angle13 of rotation step | 7.66 | 7.93 | 3.54% |
| angle14 of rotation step | 7.66 | 7.95 | 3.80% |
| angle15 of rotation step | 7.66 | 8.08 | 5.47% |
| angle16 of rotation step | 7.66 | 7.80 | 1.85% |
| angle17 of rotation step | 7.66 | 7.57 | -1.10% |
| Error (RMSE) | | 0.26 | |

Table 4.1: RMSE of estimated angles of rotation step from Middlebury temple (in degrees)

### 4.3.3 Estimation of Camera Parameters and Poses

The output of the proposed method contains the cameras' poses (rotation matrices and translations) and parameters (intrinsic matrices, radial distortion parameters).

The object's rotation matrix can be also expressed with Euler angles (Diebel, 2006), which represent three angles of a moving object's rotation about x-y-z axis in 3-D coordinate system. Fig. 4.15 illustrates the expression of Euler angles of 3-D model rotation about x-y-z axes in MeshLab software(MeshLab,

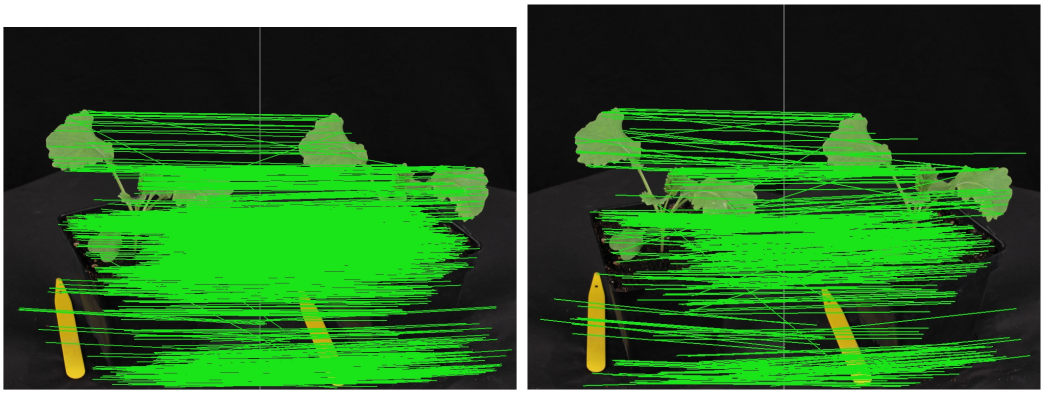
| Item | Ground truth | Estimated value | Relative error(%) |
|---|---|---|---|
| angle1 of rotation step | 7.83 | 7.93 | 1.30% |
| angle2 of rotation step | 7.83 | 8.28 | 5.81% |
| angle3 of rotation step | 7.83 | 7.80 | -0.32% |
| angle4 of rotation step | 7.83 | 7.93 | 1.36% |
| angle5 of rotation step | 7.83 | 7.79 | -0.46% |
| angle6 of rotation step | 7.83 | 7.81 | -0.22% |
| angle7 of rotation step | 7.83 | 7.67 | -1.93% |
| angle8 of rotation step | 7.83 | 7.61 | -2.74% |
| angle9 of rotation step | 7.83 | 7.49 | -4.25% |
| angle10 of rotation step | 7.83 | 7.42 | -5.19% |
| angle11 of rotation step | 7.83 | 7.44 | -4.99% |
| angle12 of rotation step | 7.83 | 7.76 | -0.88% |
| angle13 of rotation step | 7.83 | 7.37 | -5.78% |
| angle14 of rotation step | 7.83 | 7.77 | -0.66% |
| angle15 of rotation step | 7.83 | 7.60 | -2.94% |
| angle16 of rotation step | 7.83 | 7.59 | -3.06% |
| angle17 of rotation step | 7.83 | 7.87 | 0.60% |
| angle18 of rotation step | 7.83 | 7.90 | 0.96% |
| angle19 of rotation step | 7.83 | 7.87 | 0.58% |
| angle20 of rotation step | 7.83 | 8.03 | 2.65% |
| angle21 of rotation step | 7.83 | 7.82 | -0.04% |
| Error (RMSE) | | 0.23 | |

Table 4.2: RMSE of estimated angles of rotation step from Middlebury dino (in degrees)

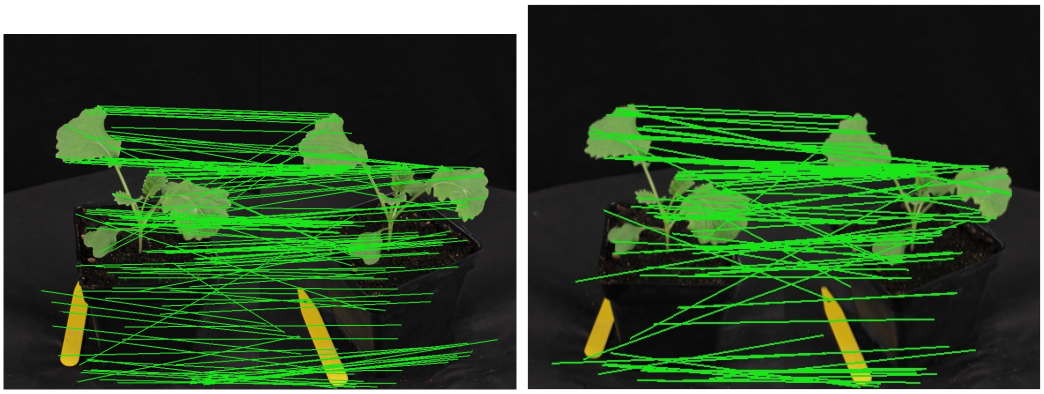
2014), where angle $\alpha$ indicates the rotation about x axis, angle $\beta$ indicates the rotation about y axis, and angle $\gamma$ indicates the rotation about z axis.

In this experiment, Euler angles were computed from the rotation matrices of 14 consecutive cameras' poses (41st-54th), which were estimated in SFM

(a) 5 consecutive images selected from 54 images of a Brassica. left to right: 48th, 49th, 50th, 51st, 52nd.



(b) left: 1029 matching between 49th-48th; right: 359 matching between 50th-48th.



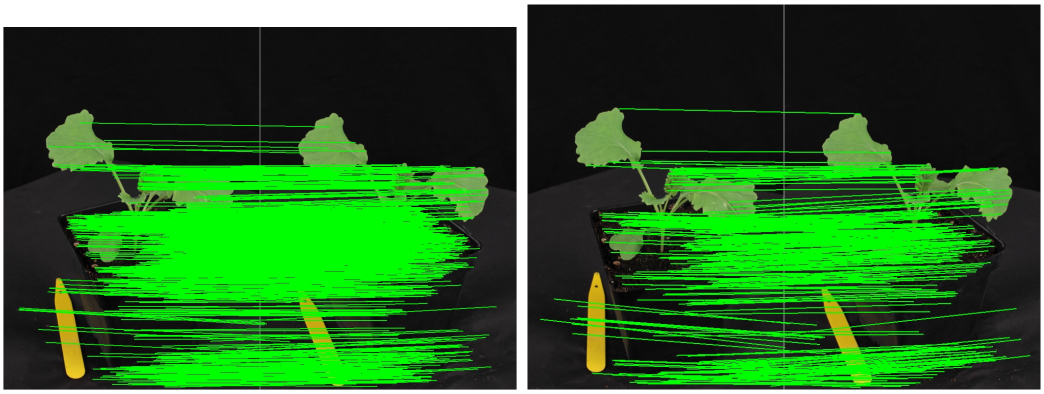
(c) left: 148 matching between 51st-48th; right: 88 matching between 52nd-48th.

Figure 4.12: SIFT feature detection and matching on 5 consecutive images of a Brassic plant.

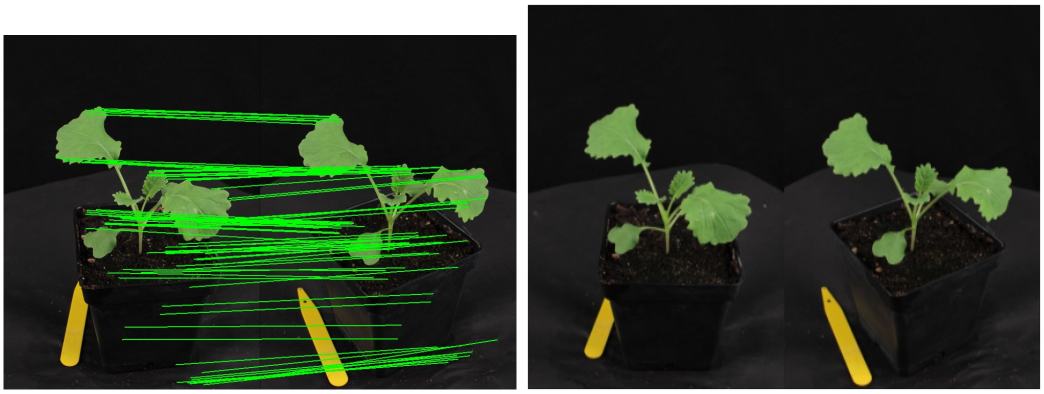
from 54 images of a Brassica. The results in Fig. 4.16 show the step of the camera rotation is not equal, but very close. The average step of rotation can be calculated − it is 8.4° in this experiment.

Due to the lack of ground truth of camera motion, I did not evaluate the

(a) 5 consecutive images selected from 54 images of a Brassic. left to right: 48th, 49th, 50th, 51st, 52nd.



(b) left: 912 inliers matching between 49th-48th; right: 290 inliers matching between 50th-48th.

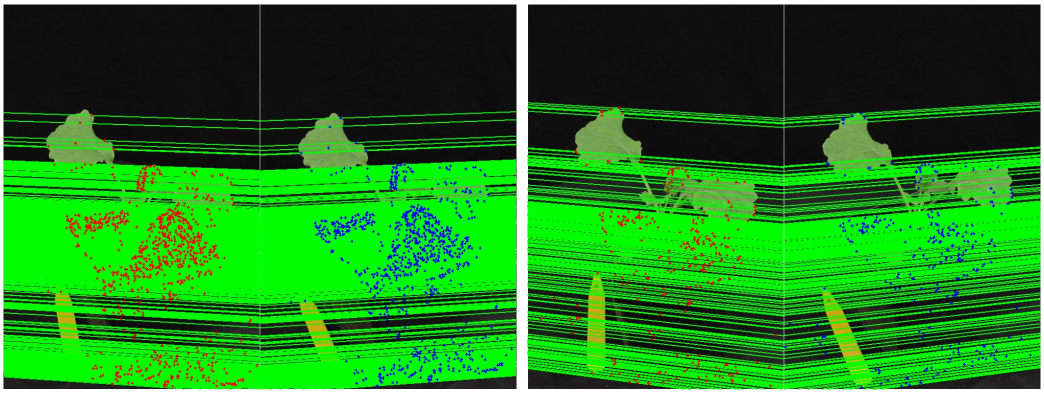


(c) left: 97 inliers matching between 51st-48th; right: 0 inliers matching between 52nd-48th

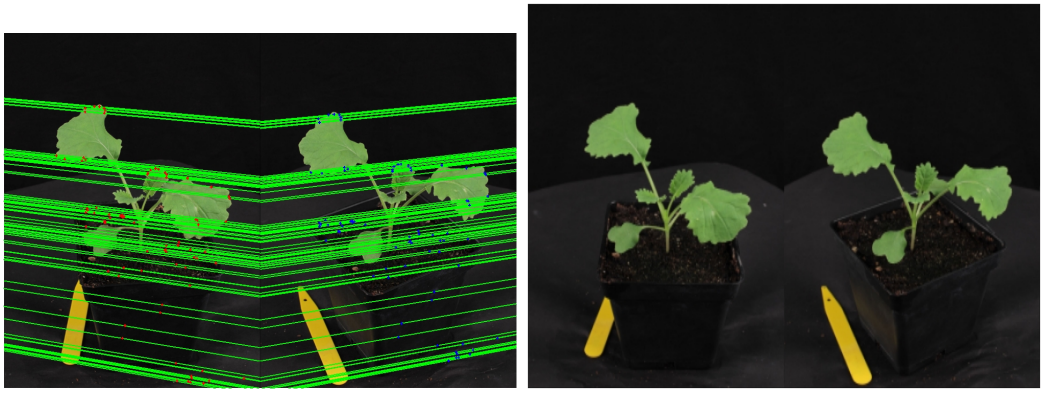
Figure 4.13: SIFT feature detection and inliers matching on 5 consecutive images of a Brassica plant.

accuracy of estimation of the cameras' poses. The cameras' poses are illustrated from side-view (Fig. 4.17) and top-view (Fig. 4.18), and approximately form a circle.

(a) 5 consecutive images selected from 54 images of a Brassica. left to right: 48th, 49th, 50th, 51st, 52nd.



(b) left: epipolar geometry between 49th-48th; right: epipolar geometry between 50th-48th.



(c) left: epipolar geometry between 51st-48th; right: epipolar geometry between 52nd-48th.

Figure 4.14: The epipolar geometry estimated from 5 consecutive images of a Brassica plant.

### 4.3.4 Comparison of Different Features on Detection and Matching

Different local invariant features can be used as alternative to SIFT in feature detection and matching. This experiment compares the performance of DAISY,

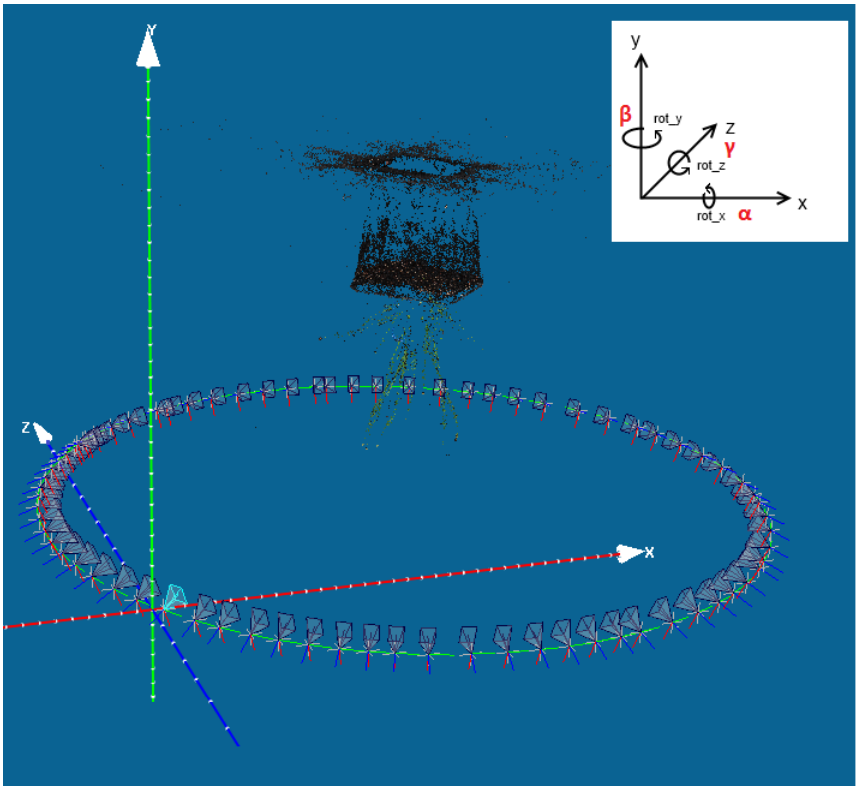


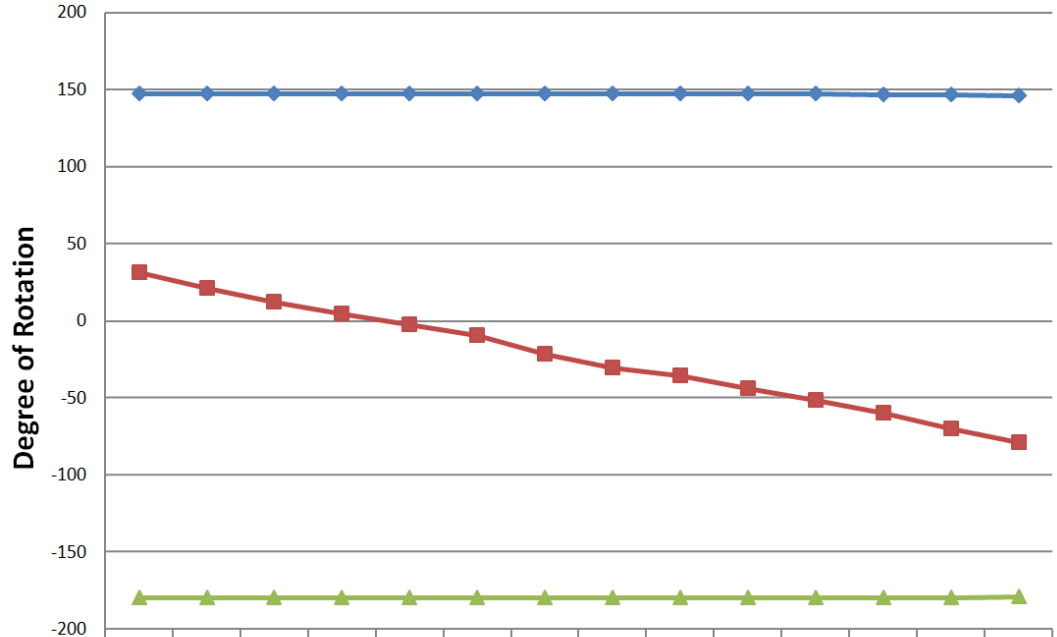

Figure 4.15: Illustration of Euler angles of 3-D model rotation about x, y, z axis in MeshLab.

AKAZE, SURF and SIFT in feature detection and 3-D points production for a Brassica and an *Arabidopsis* plant. Fig. 4.19 and Fig. 4.20 show the number of 3-D points produced by different features. The experimental results show that:

(1) DAISY and AKAZE can produce many more 3-D points than SURF, but 3-D points produced by DAISY are mainly located in planar regions like the leaves of the plant or ruler, and 3-D points detected by AKAZE focused on the boundary or edges of the plant or ruler.

(2) DAISY and AKAZE can better estimate the camera's pose than SURF in both the case of wide baseline where the rotation angle is bigger (in other words, the number of captured images is smaller) and the case of texture-less plants.

Table. 4.3 and Table. 4.4 show the performance of key points produced by different feature (DAISY, AKAZE, SURF and SIFT) detection and matching on 54 images of a Brassica and 75 images of an *Arabidopsis* respectively.




| Camera | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| α | 147.7 | 147.7 | 147.6 | 147.6 | 147.6 | 147.5 | 147.5 | 147.5 | 147.4 | 147.4 | 147.3 | 147.1 | 146.9 | 146.4 |
| β | 31.52 | 21.58 | 12.36 | 5.091 | -1.95 | -9.40 | -21.4 | -30.1 | -35.5 | -43.6 | -51.3 | -59.6 | -69.4 | -78.5 |
| γ | -179. | -179. | -179. | -179. | -179. | -179. | -179. | -179. | -179. | -179. | -179. | -179. | -179. | -178. |

Figure 4.16: Euler angles of 14 consecutive cameras' poses estimated from 54 images of a Brassica.

Figure 4.17: Illustration of the cameras' poses from side-view.

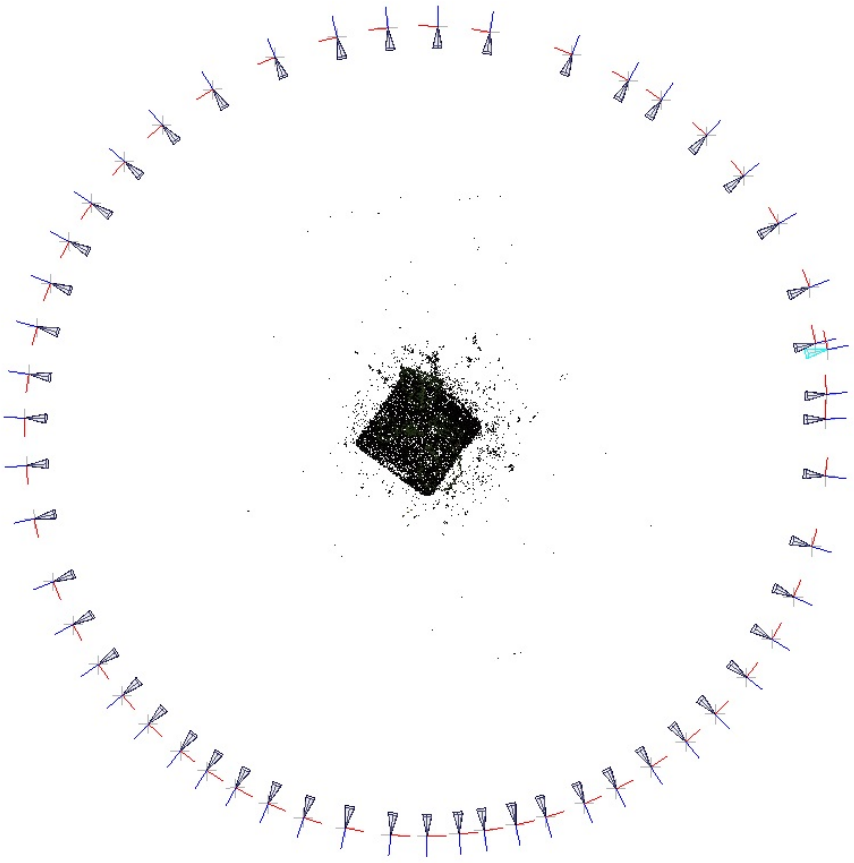
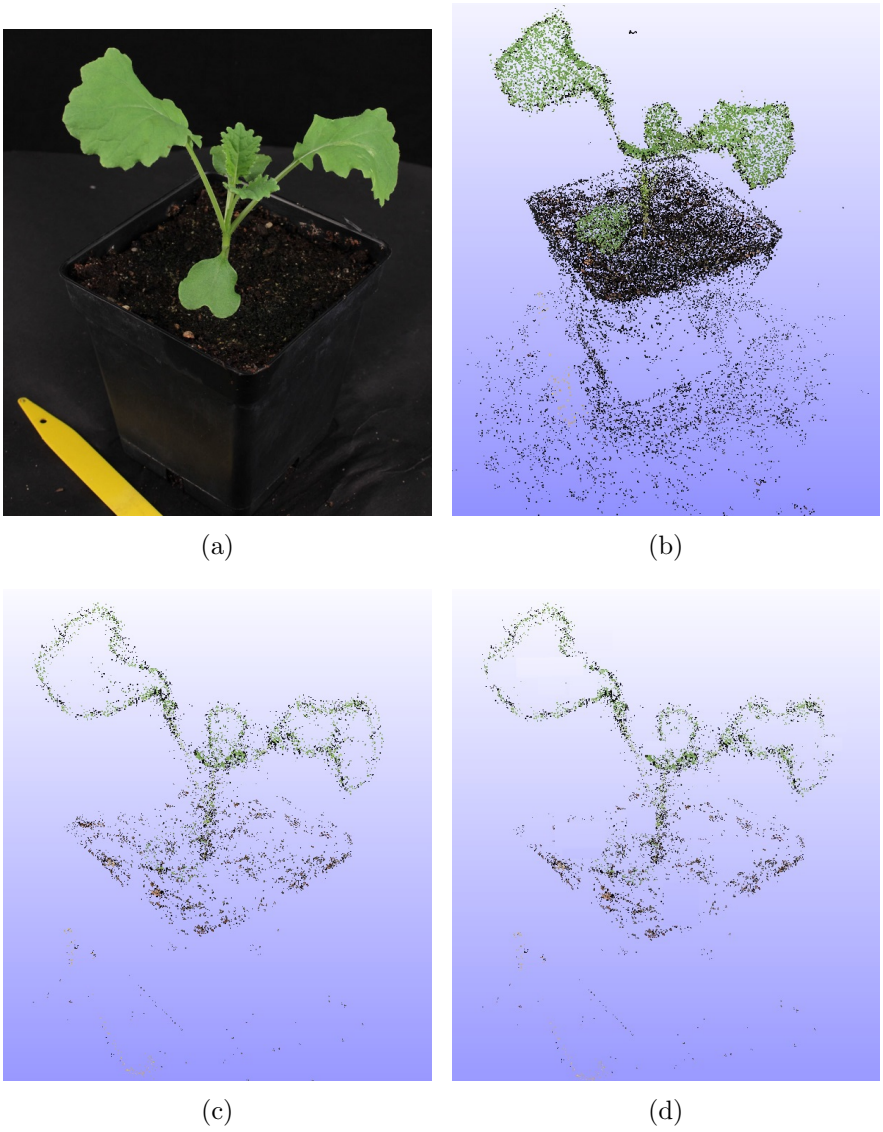


Figure 4.18: Illustration of the cameras' poses from top-view.

These experimental results show that SIFT has the best accuracy (minimal re-projection error) but the lowest inliers ratio, AKAZE has the best inliers ratio and better accuracy, and DAISY can produce the most 3-D points but has the lowest accuracy.

Generally, the computational running time for camera parameters estimation was 10∼30 minutes depending on the number of images and the number of key points of each image.

In addition, the experimental results show that rotation steps of 3°∼6° increments are optimal and feature matching does not perform well with rotation steps of more than 9°∼10° degrees in most experiments. This maximal rotation steps (9°∼10°) is lower than the maximal viewpoint changes (25°∼30°) that was found in (Moreels & Perona, 2007), where authors used digital camera and



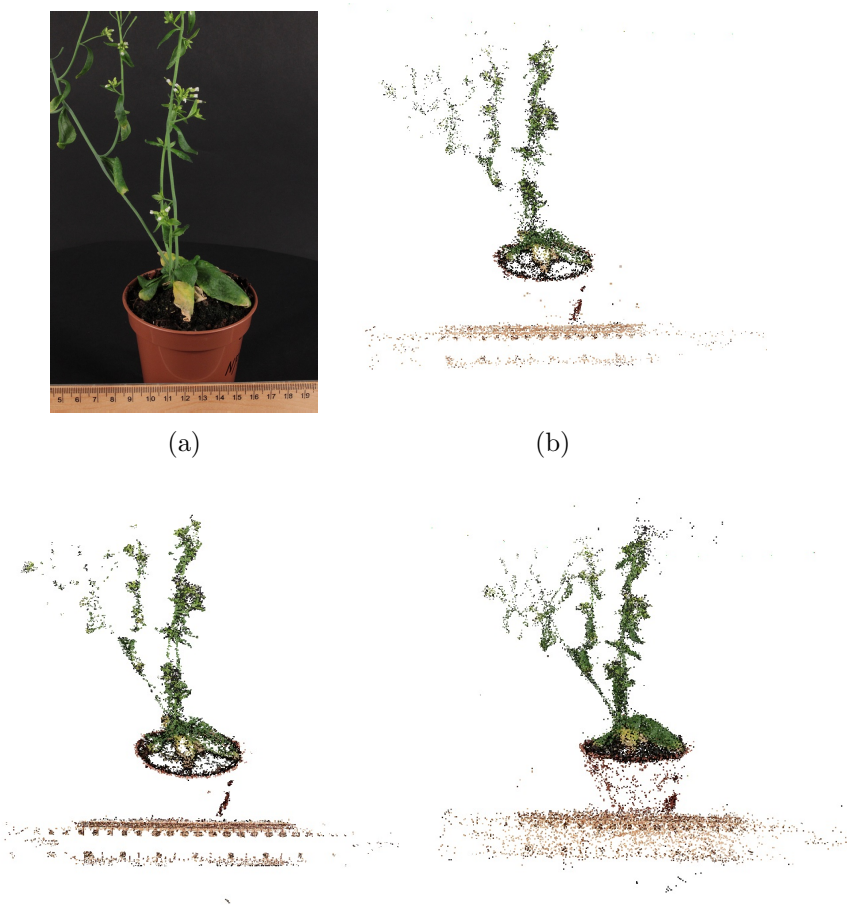
(a) (b)

(c) (d)

Figure 4.19: 3-D points produced by different features from 54 images of a Brassica. (a) sample image; (b) 48719 3-D points by DAISY ; (c) 16345 3-D points by AKAZE and (d) 8531 3-D points by SURF.

turntable to build a photo database of 100 different small man-made objects (toy, bottle, box, etc.) and evaluated the performance of a number of popular feature detectors/descriptors (e.g. SIFT, Harris, PCA-SIFT) in matching feature points under variable viewpoint and lighting conditions.

Main limitations of the proposed method is: i) the camera pose estimation can fail if the plant leaves are quite smooth and texture-less; ii) computational cost is high especially for large scale plants.



(a) (b) (c) (d)

Figure 4.20: 3-D points produced by different features from 75 images of an *Arabidopsis*. (a) sample image; (b) 16340 3-D points by SIFT. (c) 56784 3-D points by AKAZE. (d) 33335 3-D points by DAISY.

| Item | DAISY | AKAZE | SURF | SIFTGPU |
|---|---|---|---|---|
| Number of features (mean) | 7112 | 2839 | 2713 | 3610 |
| Inlier ratio (mean) | 23% | 31% | 20% | 18% |
| Reprojection Error | 0.371 | 0.256 | 0.331 | 0.242 |

Table 4.3: Comparison of key points produced by different features detection and matching on 54 images of a Brassica.

| Item | DAISY | AKAZE | SURF | SIFTGPU |
|---|---|---|---|---|
| Number of features (mean) | 6026 | 3460 | 2942 | 3255 |
| Inlier ratio (mean) | 24% | 28% | 19% | 20% |
| Reprojection Error | 0.392 | 0.295 | 0.324 | 0.283 |

Table 4.4: Comparison of key points produced by different features detection and matching on 75 images of an *Arabidopsis*.

## 4.4 Summary

This chapter proposes a camera parameters and pose estimation method based on an incremental Structure from Motion (SFM) approach. Experimental results show that the proposed method is reliable and efficient for various plants, and other features (i.e. DAISY, AKAZE) can be used as alternative to SIFT on feature detection and matching, however, SIFT reports the best re-projection error, closely followed by AKAZE. Rotation steps of 3°∼6° increments were found to be optimal and feature matching worsened with rotation steps of more than 10° degrees in most experiments.

# Chapter 5

# Multi-View Stereo Based 3-D Reconstruction for Plant Phenotying

The existing methods of 3-D reconstruction, 3-D laser/Lidar scanner, structured-light sensor (including Kinect sensor) do not work well on plants, especially on complex or even marginally occluded specimens or tiny plants. This chapter proposes an efficient and accurate image-based 3-D reconstruction system that can cope with a diversity of plant shape and size, while using equipment available to most biology labs.

## 5.1 Introduction

In recent years, Multi-view stereo (MVS) has received more and more attention, and much progress has been achieved since the Middlebury evaluation was provided Seitz et al (Seitz *et al.*, 2006a). According to the taxonomy of (Seitz *et al.*, 2006a), MVS algorithms can be divided into four categories: 3D volumetric (Vogiatzis *et al.*, 2007; Tran & Davis, 2006; Sinha *et al.*, 2007; Kolmogorov & Zabih, 2001), surface evolution (Zaharescu *et al.*, 2007; Kolev *et al.*, 2009; Delaunoy *et al.*, 2008; Pons *et al.*, 2005; Esteban & Schmitt, 2004; Faugeras & Keriven, 1997), depth-map merging [(Merrell *et al.*, 2007; Zach *et al.*, 2007; Liu *et al.*, 2009; Campbell *et al.*, 2008; Goesele *et al.*, 2006; Bradley *et al.*, 2010;

Strecha *et al.*, 2006), and featured-region growing and expansion (Furukawa & Ponce, 2010; Goesele *et al.*, 2007; Habbecke & Kobbelt, 2007; Lhuillier & Quan, 2005; Jancosek & Pajdla, 2009), where the top performing methods are capable of challenging the accuracy of laser scanners.

However, most of these approaches are not suitable for dealing with complex plant architecture. In particular, algorithms that partially rely on shape-from-silhouette techniques (Esteban & Schmitt, 2004; Vogiatzis *et al.*, 2007; Kolev & Cremers, 2008), under the assumption that a single object is visible and can be segmented from the background, are not applicable to plant images due to unavailability of fully visible silhouettes caused by occlusions.

As mentioned in previous chapters, although commencing as a relatively simple structure, the plant body plan rapidly becomes more complex due to a variety of processes – re-iterative organ formation, changes in organ spacing and identity, branching – that lead to overlapping and variable 3-D organization. Recording and measuring this complexity in a dynamic (non destructive) manner remains a serious challenge in biology.

This challenge is illustrated by the diversity of proposed methods that include laser scanning, digital camera photography, and structured light ranging. Stereo vision was used to reconstruct the 3-D surface of maize for measurement and analysis (Ivanov *et al.*, 1995). Kaminumaet *et al.*, (Kaminuma *et al.*, 2004) applied a laser range finder to reconstruct 3-D models that represented the leaves and petioles as polygonal meshes and then quantified morphological traits from these models. Biskup designed a stereo vision system with two cameras to build 3-D models of soybean plant foliage to analyze the angle of inclination of the leaves and its movement over time (Biskup *et al.*, 2007). Quan proposed a method to interactively create a 3-D model of the foliage by combining clustering, image segmentation and polygonal models (Quan *et al.*, 2006). 3-D plant analysis based on a mesh processing technique was presented in (Paproki *et al.*, 2012), where the authors created a 3-D model of cotton plant from high-resolution images using a commercial 3-D digitization product named 3DSOM. Thiago developed an image-based 3-D digitizing method for plant architecture analysis and phenotyping (Santos & Oliveira, 2012), and showed that the state-of-the-art SFM and multi-view stereovision are able to

produce accurate 3-D models with a few limitations. Li *et al.* (Li *et al.*, 2013) presented a framework to track and detect plant growth by a forward-backward 3-D point cloud analysis, where the 3-D point cloud was produced based on a structured light scanner over time. Yamazaki *et al.* presented a practical shape-from-silhouettes approach to acquire 3-D models of intricate objects with severe self-occlusions, repeated thin structures, and surface discontinuities, including tree branches, bicycles and insects (Yamazaki *et al.*, 2009).

The Patch-based Multi-view Stereo (PMVS) algorithm proposed in (Furukawa & Ponce, 2010) is considered the state of the art dense 3-D reconstruction. Based on the output of SFM, PMVS can produce dense 3-D point clouds and performs well on man-made objects and scenes. However, for plants, PMVS show significant errors caused by severe occlusions and existence of texture-less regions.

## 5.2 The Proposed Method

In this chapter, an accurate MVS 3-D reconstruction method is proposed using multi-view images of plants that takes both accuracy and efficiency into account.

Briefly, the system consists of three steps: SFM, stereo matching, and depth computation and merging. The proposed method directly selects the starting pair of images to begin for an incremental SFM to estimate the camera parameters as described in the previous chapters, and then rectifies the stereo pairs, computes and refines depth-maps. Finally, all the refined depth-maps are merged together into a single global 3-D point cloud to provide a final reconstruction. The framework is illustrated as Fig. 5.1.

### 5.2.1 Stereo Pair Selection

Not all image pairs are eligible for stereo matching. The selection of stereo image pairs is important not only for the accuracy of the final MVS result but also for the time performance of the system. A good candidate neighboring image pairs should have sufficient ray intersection angles with the reference images, and have a suitable baseline neither too short to degrade the reconstruction ac-

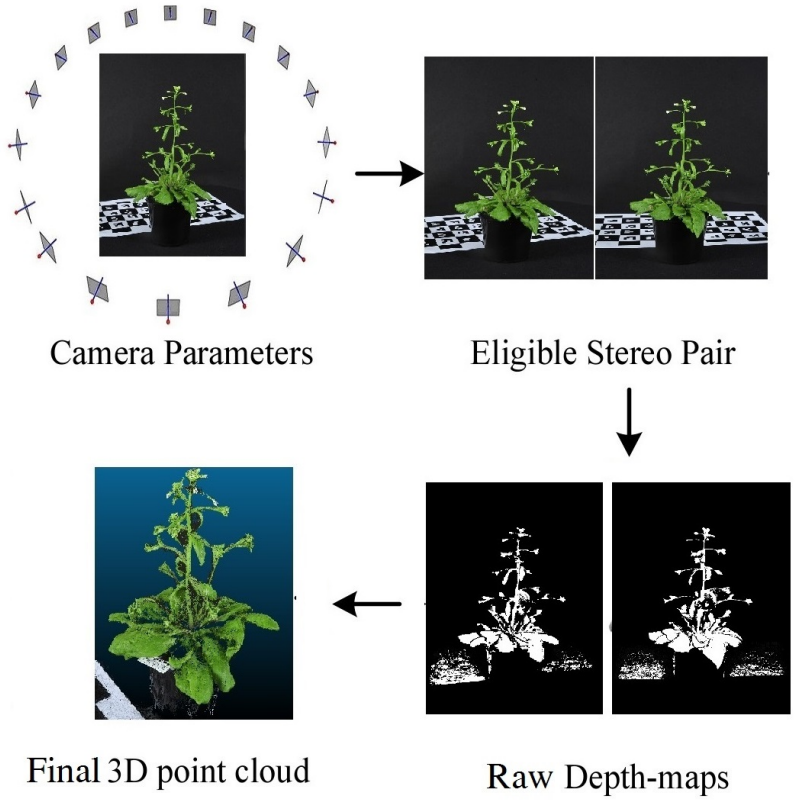



Figure 5.1: The framework of the proposed 3-D reconstruction method, including three main steps: SFM, stereo matching, and depth computation and merging.

curacy nor too long to have less common coverage of the scene. Hence, the pair selection is based on two statistics from the $i$-th possible image pair: the angle between principal view directions and the distance between camera optical centers.

To select eligible stereo pairs, I assumed that the sequence of images has $n$ frames, and for the $i$-th one, I computed $\theta_{ij}$ ( $j = 1, 2, ..., n$), the angle between principal view directions of cameras $i$ and $j$, and $d_{ij}$ $(j = 1, 2, ..., n)$ , the distance between optical centers of cameras $i$ and $j$ (see Fig. 5.2). In the system, eligible stereo pairs are determined by the following rules:

i) First, according to the experimental results that rotation steps of 3°∼6° increments are optimal and feature matching does not perform well with rotation steps of more than 9°∼10° degrees in most experiments (see Chapter 4), I

choose the stereo pairs whose $\theta_{ij}$ satisfies $3° \leq \theta_{ij} \leq 10°$.

ii) Second, I compute the median $\bar{d}$ of $d_{ij}$, and remove the stereo pair whose $d_{ij} \geq 2\bar{d}$ or $d_{ij} \leq 0.05\bar{d}$.

Finally, it can be seen that each remaining image view $i$ has at most two stereo pairs with its neighbors $j$ and the number of eligible stereo pairs is fewer than $2n$ based on the above rules.

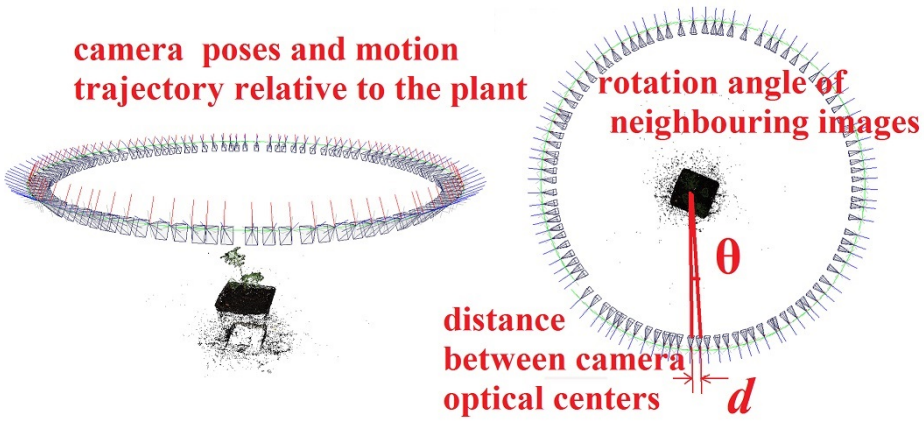



Figure 5.2: Select eligible stereo pairs from a sequence of images. The $\theta$ is the angle between principal view directions of two cameras. The $d$ is the distance between optical centers of two cameras.

### 5.2.2 Depth Maps Computation

Let $I_b$ and $I_m$ be a rectified eligible image pair, my objective is to estimate a set of depth maps. The classic SGM (Semi-Global Matching) algorithm (Hirschmuller, 2008) aims to recover disparities across stereo pairs by minimizing the following global cost function:

$$E(\mathbf{D}) = \sum_{\mathbf{x}_b}(C(\mathbf{x}_b, \mathbf{D}(\mathbf{x}_b)) + \sum_{\mathbf{x}_N} P_1 T[\|\mathbf{D}(\mathbf{x}_b) - \mathbf{D}(\mathbf{x}_N)\| = 1] + \sum_{\mathbf{x}_N} P_2 T[\|\mathbf{D}(\mathbf{x}_b) - \mathbf{D}(\mathbf{x}_N)\| > 1])$$

Where $\mathbf{D}(\mathbf{x}_b)$ represents the disparity estimations of all base image pixels $\mathbf{x}_b$ of $I_b$, $\mathbf{x}_N$ denotes base image pixels in the neighborhood of $\mathbf{x}_b$, $P_1$ and $P_2$ are penalty constants to control the gain of a little or a larger disparity changes

respectively, $T$ is an operator evaluating to one if the subsequent condition is true and evaluate to zero else. $C(\mathbf{x}_b, \mathbf{D}(\mathbf{x}_b))$ computes the pixel-wise similarity measures.

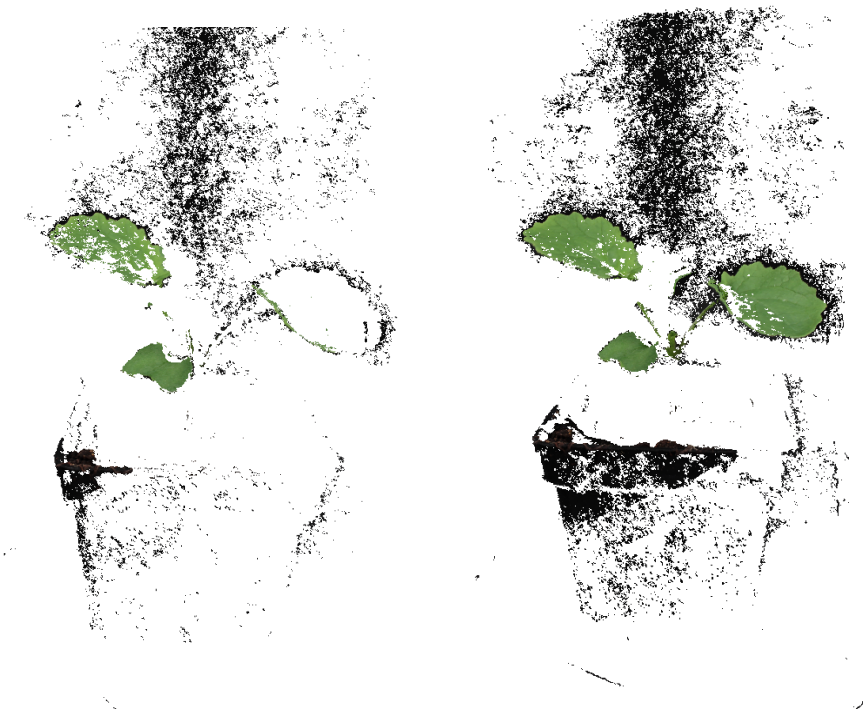
Our method is based on the OpenCV API which implements a memory and time efficient modification of SGM, where a hierarchical approach was proposed to initialize and refine the Mutual Information (MI) matching cost, and initial disparity images were computed by matching high level (low resolution) image pyramids. The resulting disparities were then used to refine the MI matching cost for processing the subsequent pyramid level. See OpenCV API for details.

### 5.2.3 Depth Maps Merging

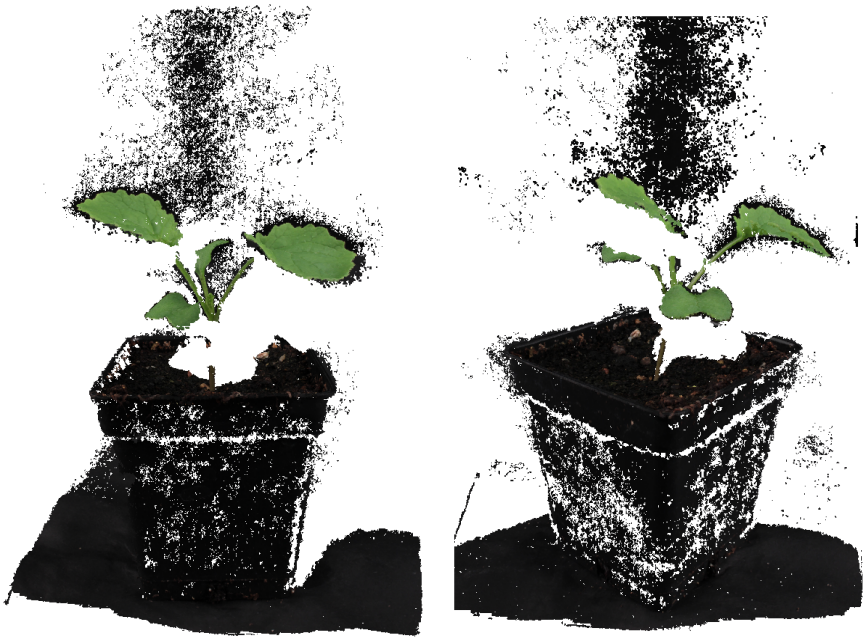
Because the raw depth maps are view-dependent and may not completely agree with each other on common areas due to matching errors or occlusions, a refinement process is necessary to enforce consistency over neighboring images. For each point $p$ in the depth map of image $I_t$, $p$ is firstly back-projected to a 3-D point $X$ using its depth value and the camera parameters (Hartley & Zisserman, 2004), and then checked geometric consistency by comparing the deviations of depth values computed at the projections of $X$ in neighboring images. After removing those in-consistent points, all the depth-maps had been merged into a single dense 3-D point cloud. While the depth-maps may contain lots of redundancies, because different depth-maps may have common coverage of the object scene in neighboring images, I do not consider removing the redundancies and noises at this stage for the sake of cost-effectiveness.

Fig. 5.3 illustrates the procedure of the depth maps to be merged incrementally. It shows that there are lots of noisy points (mainly black or dark points, that caused by the background) in the 3-D point cloud. These black or dark noisy points can be removed from the final 3-D point cloud using a simple color filtering where the (R, G, B) of color components of a noisy point usually satisfies ($R \leq 50$, $G \leq 70$, $B \leq 50$). Although those 3-D points produced by the black/dark parts of a plant (including black plant pot or dark soils) could also be removed from the final 3-D point cloud, the simple color filtering still worked efficiently in our experiments due to the fact that most plants hardly had the black or dark organs which appearance color was approximate to the

black background curtain.



(a) Left: 1st depth map. Right: 5th depth map.



(b) Left: 15th depth map. Right: 20th depth map.

Figure 5.3: Illustrations of the procedure of the depth maps to be merged incrementally.

## 5.3 Experimental Results

The experiments were run on an Intel i7 laptop, Dell Precision M6700 with 16G RAM and Nvidia graphics card. In the comparison analysis experiments, the available 3-D reconstruction methods, such as PMVS (Furukawa & Ponce, 2010), CMPMVS (Jancosek & Pajdla, 2011), Kinect sensor (ReconstructMe) (Kinect, 2011)(ReconstructMe, 2014), and commercial 3-D modeling techniques, 3DSOM (3DSOM, 2014) and Artec Eva 3-D scanner (Artec, 2013), used their default parameters.

Compared to the state-of-art MVS methods, such as PMVS(Furukawa & Ponce, 2010), CMPMVS(Jancosek & Pajdla, 2011), the experimental results show that the proposed method can produce much more dense and detail 3-D point clouds from multi-view images of plants. Fig. 5.4 shows two different genotype brassicas. Fig. 5.5 shows the 3-D reconstructions of two brassicas. It can be seen that the 3-D reconstructions is capable to represent the brassica with a complete 3-D shape and the detailed organs (i.e. stems, leaves).

Figure 5.4: Sample image from two different species of brassica, which have significant different in leaf shape.

Fig. 5.6 shows the 3-D reconstruction point clouds of diverse plants produced by the proposed method. Fig. 5.7 shows the 3-D reconstruction point clouds of

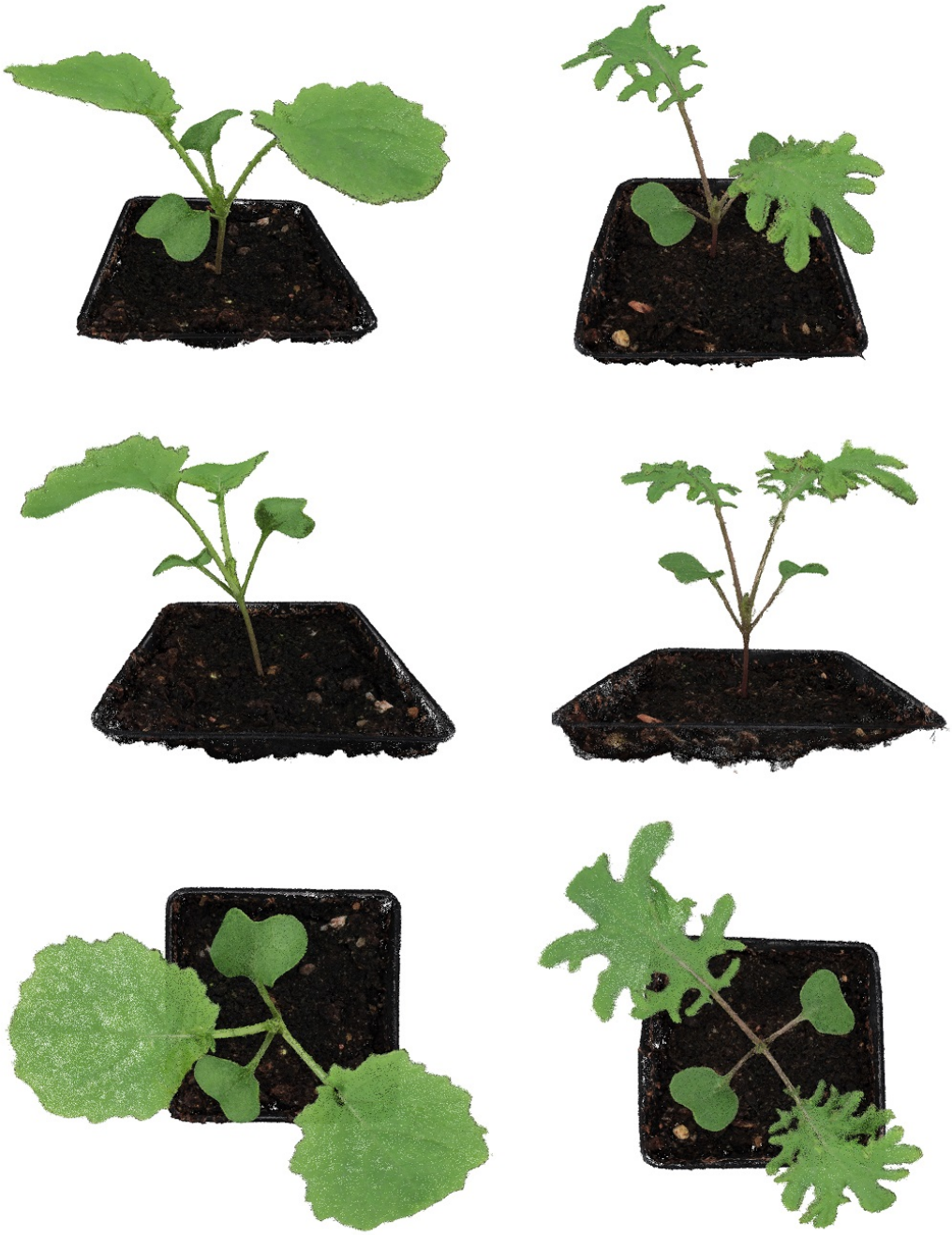


Figure 5.5: Illustration of 3-D reconstructions of two different species of Brassica, from front-view (top), side-view (middle) and top-view (bottom) respectively. The 3-D reconstruction point clouds represented the complete shapes and detailed surface textures of two Brassica.

diverse *Arabidopsis* strains and mutants produced by the proposed method.

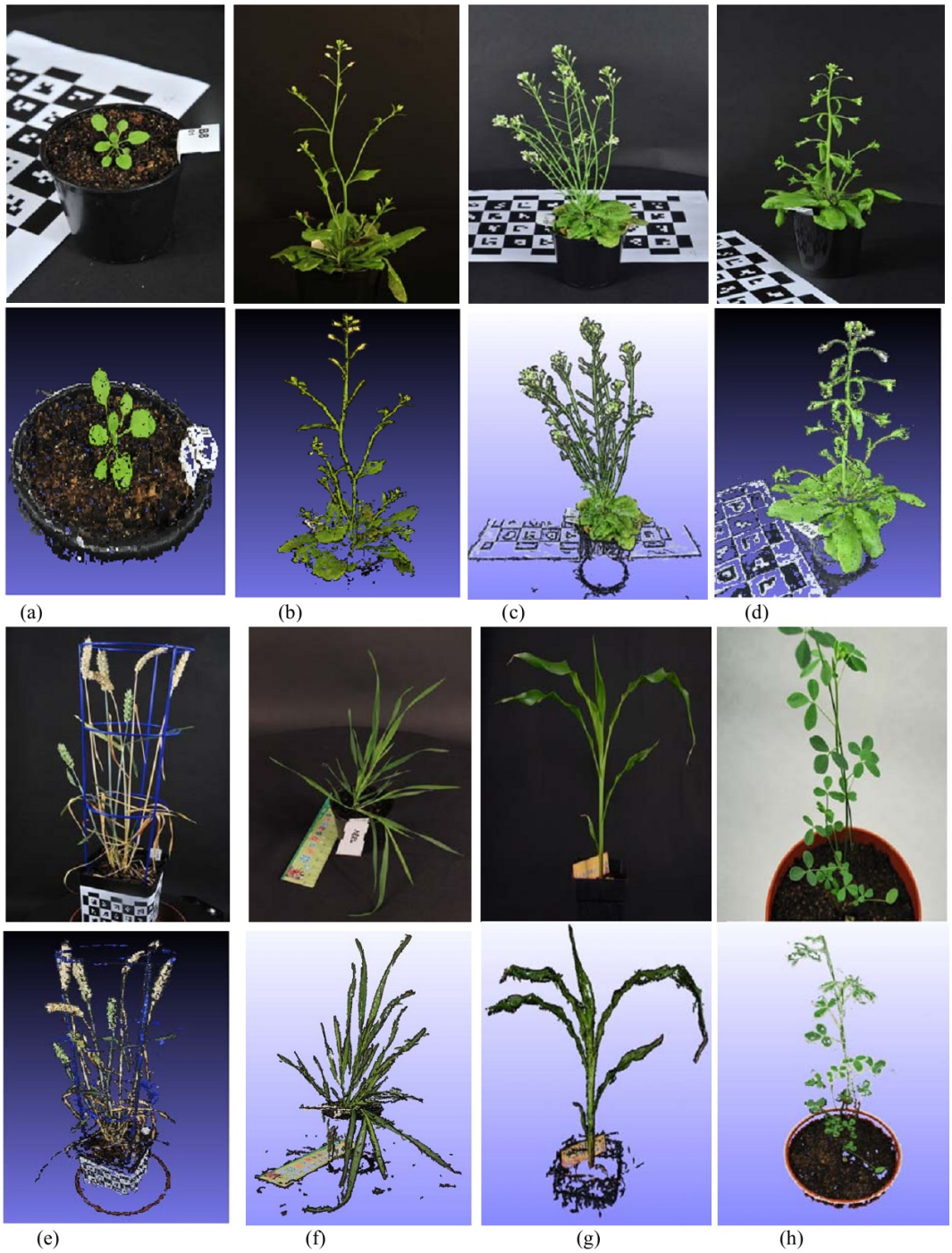



Figure 5.6: Raw images (odd rows) and 3-D reconstructions (even rows) of diverse plants. (a)-(d) Arabidopsis strains and mutants; (e) wheat; (f) Brachypodium; (g) maize and (h) clover.

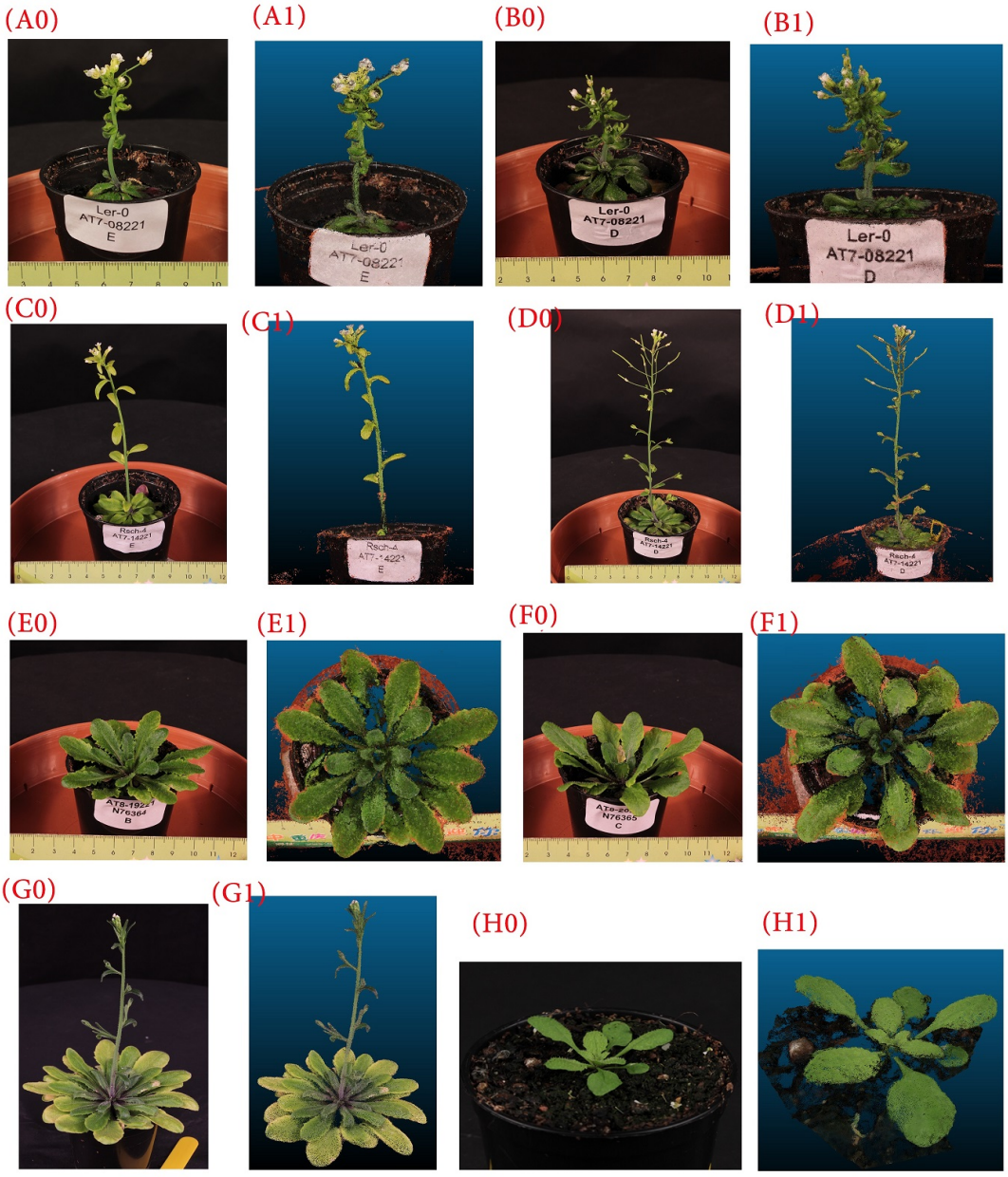



Figure 5.7: Raw images (odd columns) and 3-D reconstruction point clouds (even columns) of diverse *Arabidopsis* strains and mutants. (A0)-(H0) The sample 2-D images of *Arabidopsis* strains and mutants; (A1)-(H1) 3-D reconstruction point clouds of *Arabidopsis* strains and mutants.

### 5.3.1 Quality Analysis of the 3-D Reconstruction

3-D sensors developed for non-destructive plant phenotyping are based, variously, on laser scanning, structured light, multi-view stereo, etc., and each has its

own merits and limitations. Unfortunately, there does not exist a widely accepted benchmark that can be used to compare those different methods. Therefore, I have investigated the existing methods of 3-D reconstruction and found that the 3-D laser/Lidar scanner or the structured-light scanner (including Kinect) do not work well on plants, especially for complex or even marginally occluded specimens or tiny plants.

Fig. 5.8 shows the 3-D reconstruction results from 62 images of a Physalis plant, produced by the existing state of art methods PMVS (Furukawa & Ponce, 2010), CMPMVS (Jancosek & Pajdla, 2011), commercial 3-D modeling techniques (3DSOM (3DSOM, 2014), Artec Eva 3-D scanner (Artec, 2013), Kinect sensor (Kinect, 2011) and the proposed method, respectively.

Fig. 5.9 shows the 3-D reconstruction results from 60 images of an Oilseed rape plant, produced by PlantEye (PlantEye, 2015), Artec Eva 3-D scanner and the proposed method, respectively.

In the cases of Fig. 5.8 and Fig. 5.9, the proposed method took approximately 15 minutes to produce the final 3-D point clouds up to around 5 million points respectively.

The PMVS software downsampled the 2-D images, and was therefore faster than the proposed method − it took only 3 minutes to yield about 20k points, but the quality of the 3-D point cloud was low in the sense of sparse points and holes as shown in Fig. 5.8 (A5).

The CMPMVS software was very slow, and took around 180 minutes to yield 65k points with some stems and leaves missing in the top as shown in Fig. 5.8 (A6).

Although the Artec Eva (a commercial handheld structured light scanner) is real-time 3-D scanners, they had to be held by hand carefully to scan around the plant, and usually missed small stems and leaves as shown in Fig. 5.8 (A3) and Fig. 5.9 (B3) respectively.

ReconstructMe (ReconstructMe, 2014) is a 3-D real-time scanning software that supports a wide range of commodity RGBD sensors such as the Microsoft Kinect, the PrimeSense Carmine or the ASUS Xtion. ReconstructMe performs the entire reconstruction in metric space and the result can be exported to various CAD formats such as STL, OBJ, 3DS, and PLY. Paulus *et al.*(Paulus *et al.*,

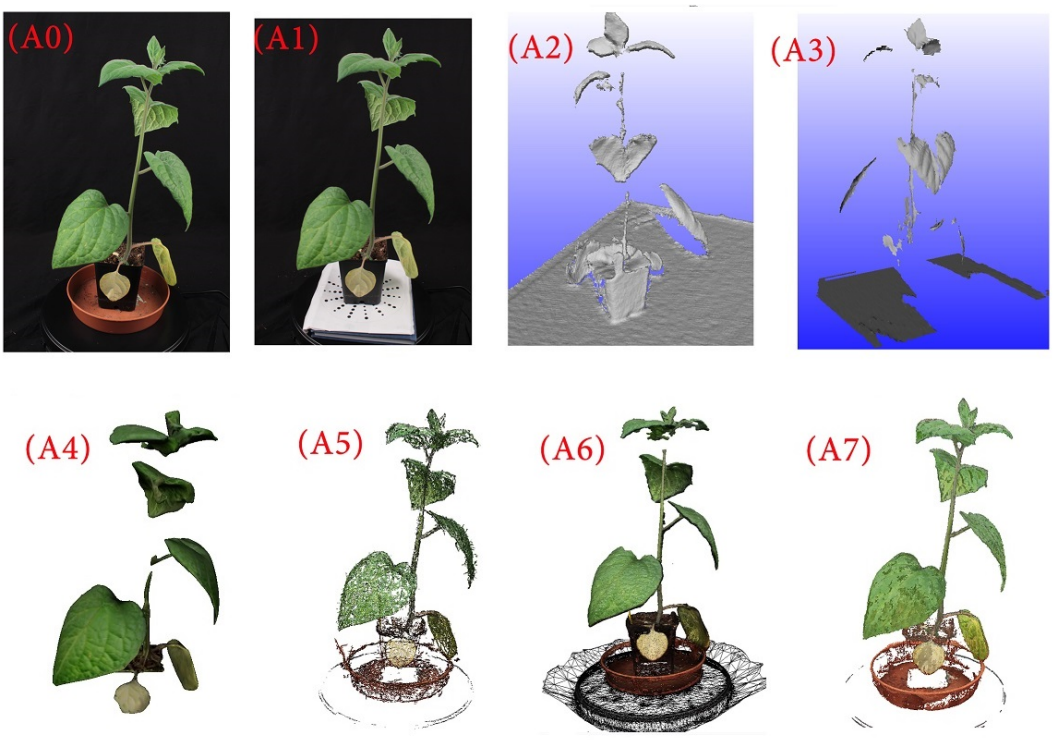



Figure 5.8: 3-D reconstruction results from 62 images of a Physalis, produced by the existing state of art methods PMVS, CMPMVS, 3DSOM, Artec Eva 3-D scanner, Kinect sensor and the proposed method, respectively. (A0) a sample image of the Physalis; (A1) a sample image of the Physalis with a marker pattern, only for 3DSOM; (A2) 3-D reconstruction result by Kinect and ReconstructMe software; (A3) 3-D reconstruction result by Artec Eva 3-D scanner; (A4) 3-D reconstruction result by 3DSOM; (A5) 3-D reconstruction result by PMVS; (A6) 3-D reconstruction result by CMPMVS; (A7) 3-D reconstruction result by the proposed method.

2014c) investigated the applicability of several low-cost 3-D imaging systems for the task of plant phenotyping. To this end, the Microsoft Kinect and the DAVID laser scanner were compared to the high-precision close-up Perceptron v5 laser scanner, which served as a reference. Experimental results in (Paulus *et al.*, 2014c; Khoshelham & Elberink, 2012) show that with increasing distance, the accuracy of Kinect depth measurement decreases from a standard deviation (SD) of a few millimeters to about 40 mm, and the point-to-point distance increases from 0.9 mm to 7mm. These results may explain why in our experiments Kinect only worked on some plants with wide leaves and always failed on detecting the main stems and branches (see Fig. 5.8(A2)). In the

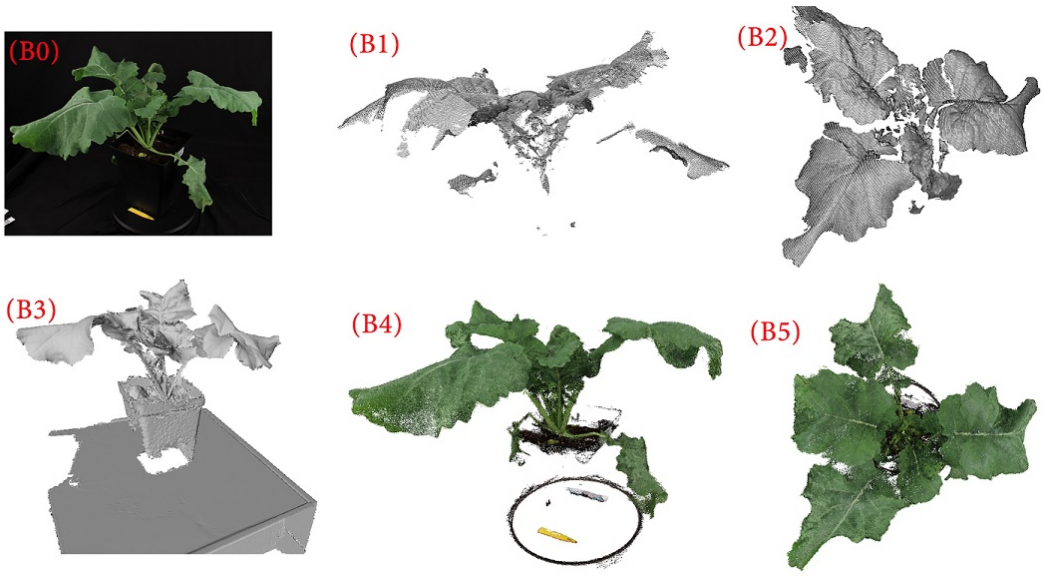



Figure 5.9: 3-D reconstruction results from 60 images of an Oilseed rape, produced by PlantEye, Artec Eva 3-D scanner and the proposed method, respectively. (B0) a sample image of an Oilseed rape; (B1) side-view of the 3-D reconstruction by PlantEye; (B2) top-view of the 3-D reconstruction by PlantEye; (B3) 3-D reconstruction result by Artec Eva 3-D scanner; (B4) side-view of the 3-D reconstruction by the proposed method; (B5) top-view of the 3-D reconstruction by the proposed method;

experiment proposed by Li *et al.* (Li *et al.*, 2015), the 3-D model of a tomato plant yielded by Kinect sensor also missed many tiny or thin parts, such as small branches and leaves (see Fig. 1.14(c)). However, the reason why reconstructed result by Kinect missed the tiny or thin parts was also likely because of the software implementation to do the registration, noise filering and fusion processes.

3DSOM is a commercial shape-from-silhouette 3-D modeling software for creating 3-D meshes from 20 multi-view images. It calculates camera parameters and poses by means of special calibration pattern (as shown in Fig. 5.8 (A1)), and then segments the object from background to get its silhouettes at different viewpoints. 3DSOM usually runs interactively and takes about 10 minutes to yield a 3-D mesh. Texture-mapping can be used to add color information to the 3-D model, as shown in Fig. 5.8(A4).

PlantEye is a commercial laser scanning system that uses a laser light section to acquire 3D point clouds. The scanner projects a very thin near infrared

laser line beneath itself onto the object of interest, meanwhile a camera that is mounted at the end of the scanner will take a picture of this line under a well defined angle. Obviously, PlantEye suffers from occlusion problem as shown in Fig. 5.9 (B1)(B2).

The experimental results show that all organs of the plants can be visually represented in the resulting 3-D point clouds reconstructed by the proposed method, including the very tiny organs, like small branches, as shown in Fig. 5.8 (A7) and Fig. 5.9 (B4)(B5).

### 5.3.2 3-D measurement of Silique Lengths of *Arabidopsis*

This experiment tested whether the point clouds produced by the proposed 3-D reconstruction method are precise. One main purpose of 3-D plant phenotyping is to measure the derivation of 3-D data of plant organs, such as plant leaves, stems, siliques, etc. In this case study, two *Arabidopsis* plants were grown in different temperatures (4°C and 12°C respectively). I firstly imaged the mature *Arabidopsis* and produced its 3-D point cloud and then manually extracted some phenotypic data such as silique lengths.

Fig. 5.10 shows measurement results of silique lengths of two *Arabidopsis*, where: (A1)-(A3) are sample images of two *Arabidopsis* grown in 4°C, cut stem in 2-D image acquisition and in flatbed scanner, respectively; (B1)-(B3) are sample images of two *Arabidopsis* grown in 12°C, cut stem in 2-D image acquisition and in flatbed scanner, respectively; (C1)-(C2) indicate the 3-D reconstruction point clouds of the cut stem in A2 by the proposed method and CT 3-D scanner respectively; (D1)-(D2) indicate the 3-D reconstruction point clouds of the cut stem in B2 by the proposed method and CT 3-D scanner respectively; (E) show the correlation of 3-D measurements of silique lengths from C1 compared to the measurements from CT 3-D scanner; (F) show the correlation of 3-D measurements of silique lengths from C1 compared to the measurements from the 2-D image.

Because the images were uncalibrated, we used a known size object as reference to calculate the real geometric dimension of the plant. In this case, the

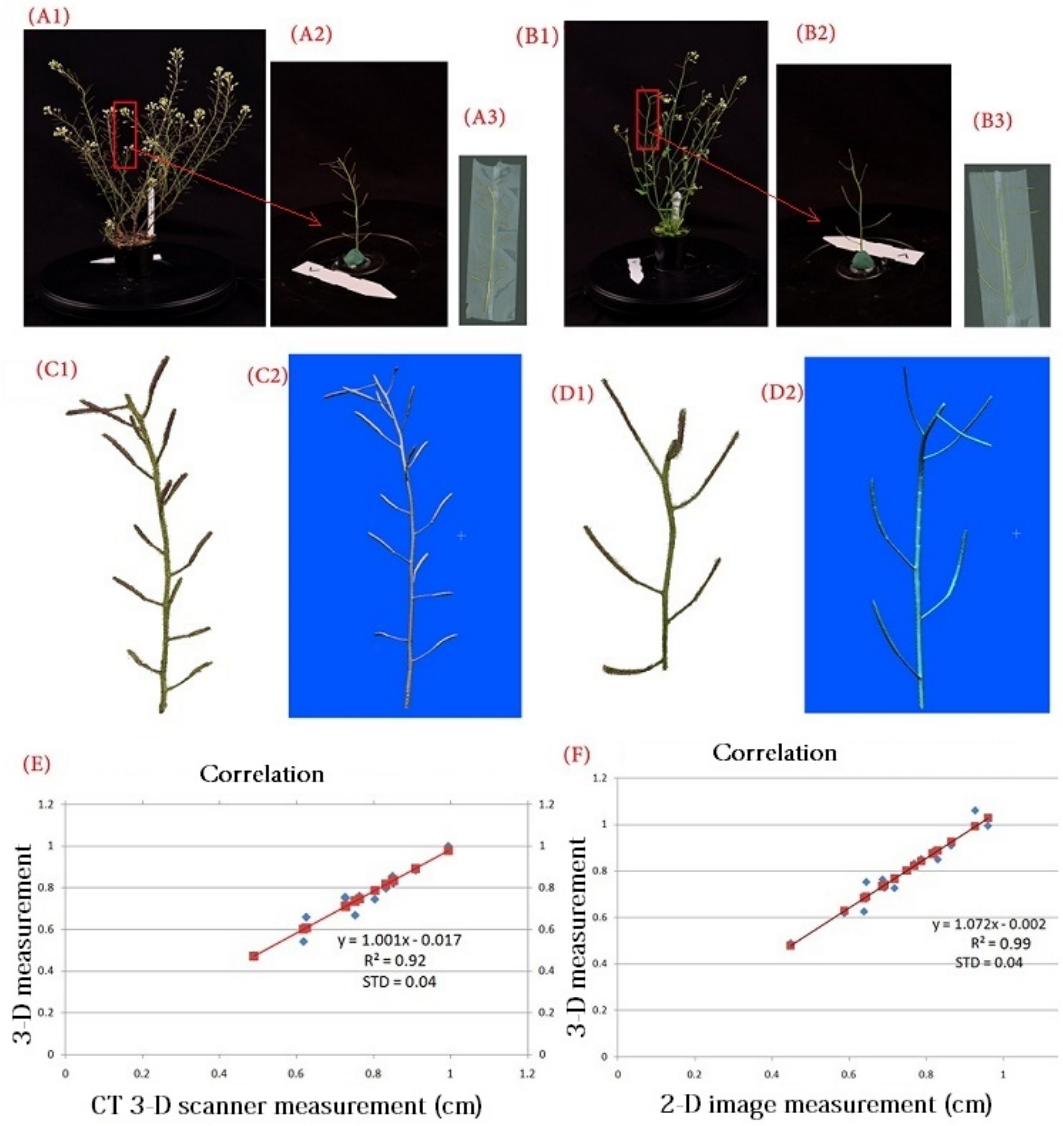



Figure 5.10: 3-D measurement results of silique lengths of two different *Arabidopsis* grown in 4°C and 12°C, respectively. (A1)-(A3) sample images of two *Arabidopsis* grown in 4°C, cut stem in 2-D image acquisition and in flatbed scanner, respectively; (B1)-(B3) sample images of two *Arabidopsis* grown in 12°C, cut stem in 2-D image acquisition and in flatbed scanner, respectively; (C1)-(C2) the 3-D reconstruction point clouds of the cut stem in A2 by the proposed method and CT 3-D scanner respectively; (D1)-(D2) the 3-D reconstruction point clouds of the cut stem in B2 by the proposed method and CT 3-D scanner respectively; (E) the correlation of 3-D measurements of silique lengths from C1 compared to the measurements from CT 3-D scanner; (F) the correlation of 3-D measurements of silique lengths from C1 compared to the measurements from the 2-D image.

white label provided a reference object and its width was 2 cm. After 3-D reconstruction of the Arabidopsis, we killed and scanned them with a CT 3-D scanner and a flatbed scanner respectively. For scanned 2-D images, we used the software *ImageJ* (ImageJ, 2014) to measure the silique lengths; For CT scanned slices, we rebuilt the 3-D models using the software *3D Slicer* (Slicer, 2014) and measured the silique lengths as ground truth.

From the measurement results of silique lengths, we can find that the 3-D reconstruction result has high correlations of $R^2 = 0.99$ with the ground truth from 2-D images and $R^2 = 0.92$ with the ground truth from CT scanning.

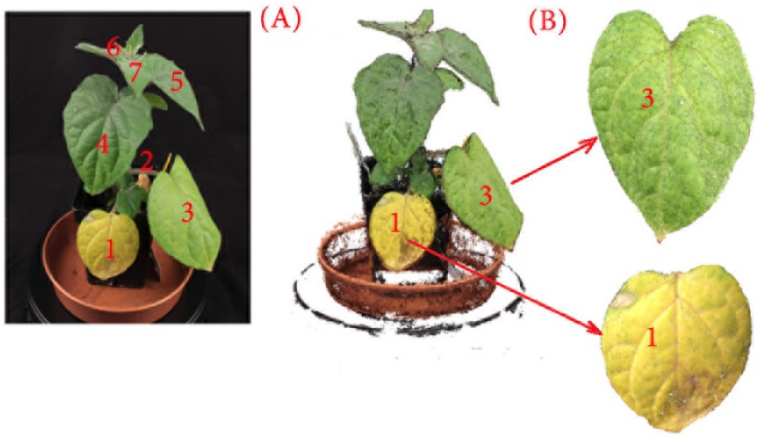
### 5.3.3 3-D Measurement of Leaf Areas

In order to extract leaf areas for plants, we captured 60 images of a growing Physalis and reconstructed its 3-D point cloud. I used MeshLab (MeshLab, 2014), an open source 3-D processing software, to build the 3-D mesh surfaces and then computed the area of each leaf. Fig. 5.11 shows the 3-D measurement results of the leaf areas of the Physalis, where (A) is sample image of the Physalis; (B) is 3-D reconstruction point cloud of the Physalis; (C) lists the measurement results and ground truth of the leaf areas; (D) shows the correlation of the 3-D measured leaf areas compared to manual measurement from the 2-D leaf images.

The ground truth of each leaf area was obtained manually from their 2-D images. The results show that the 3-D reconstruction result has a high correlation of $R^2 = 0.97$ with the ground truth and a RMSE of 5.26 $cm^2$.

### 5.3.4 Performance Analysis of Different Methods

Fig. 5.12 shows the 3-D reconstruction of an *Arabidopsis* from 66 images , produced by the existing state of the art methods (Furukawa & Ponce, 2010), (Jancosek & Pajdla, 2011) and the proposed method respectively. In the case of Fig. 5.12, the proposed method ran 15 minutes to produce the final 3-D point cloud. The CMPMVS software from (Jancosek & Pajdla, 2011) was very slow and ran around 182 minutes. The PMVS software from (Furukawa & Ponce, 2010) was faster than the proposed method and only ran 6 minutes but quality




(C)

| | Ground truth (cm²) | Area of 3D mesh (cm²) |
|---|---|---|
| leaf1 | 42.3 | 39.0 |
| leaf2 | 94.0 | 93.4 |
| leaf3 | 115.9 | 123.2 |
| leaf4 | 76.2 | 72.3 |
| leaf5 | 54.0 | 61.1 |
| leaf6 | 33.1 | 28.2 |
| leaf7 | 13.3 | 19.7 |

(D)

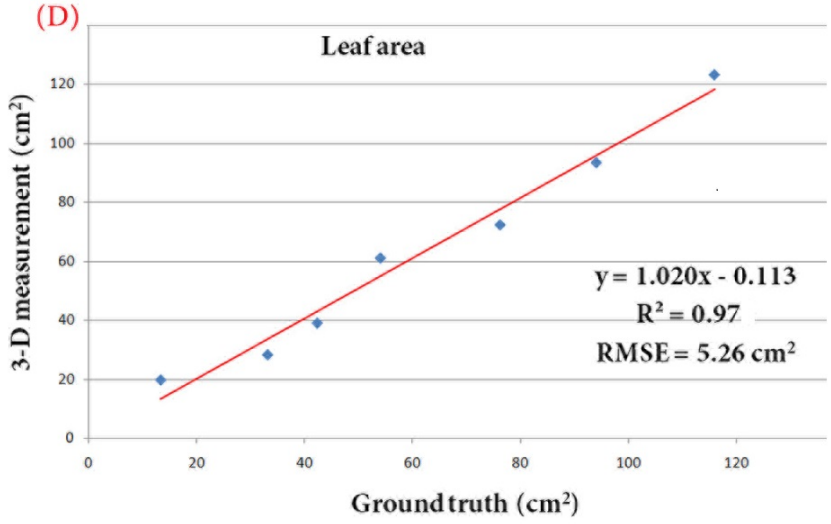



Figure 5.11: 3-D measurement results of the leaf areas of a Physalis. (A) sample image of the Physalis; (B) the 3-D reconstruction point cloud of the Physalis; (C) the measurement results and ground truth of the leaf areas; (D) the correlation of the 3-D measured leaf areas compared to manual measurement from the 2-D leaf images.

was poor. The proposed method is able to produce more dense and complete 3-D point cloud of plants than previously reported methods.

In order to evaluate the quality of the 3-D reconstruction, I have extracted the sizes of stems and leaves from resulting 3-D point cloud, and compared the results with limited ground truth that came from the manual measurement or commercial system, i.e. *WinDIAS* leaf image analysis system (WinDIAS, 2014). Fig. 5.13 shows an *Arabidopsis* and the 3-D point clouds produced by CMPMVS, PMVS and the proposed method respectively. Table. 5.1 shows the analysis of measurement errors of some leaves and main stems, where only the proposed method can produce complete and dense 3-D point clouds and extract completed geometric information.

Experimental results show that the accuracy of the proposed method is high enough to distinguish different genotypes plant and precision is stable enough to adapt slightly deviated experimental conditions.

| Item | Ground truth (mm) | Proposed method (mm) | Ref(PMVS) (mm) | Ref(CMPMVS) (mm) |
|---|---|---|---|---|
| Length of leaf1 | 21.5 | -1.2 | N/A | N/A |
| Width of leaf1 | 10.7 | 0.9 | N/A | N/A |
| Length of leaf2 | 19.5 | -0.1 | N/A | N/A |
| Width of leaf2 | 7.6 | 0.6 | N/A | N/A |
| Length of leaf14 | 41.7 | -0.5 | N/A | 2.2 |
| Width of leaf14 | 21.1 | 1.0 | N/A | 3.5 |
| Length of leaf16 | 36.7 | -1.7 | N/A | 2.6 |
| Width of leaf16 | 19.3 | -0.9 | N/A | 3.1 |
| Length of main stem | 84.8 | -2.1 | -3.4 | N/A |
| width of main stem | 1.4 | -0.1 | NA | N/A |
| Root mean square error (RMSE) | | 1.1 | | |

Table 5.1: Measurement errors of the 3-D reconstructions of an *Arabidopsis*.

## 5.4 Summary

This chapter demonstrates that 3-D reconstruction based multi-view images can be employed as a low-cost, powerful alternative for non-destructive plant phenotyping. The proposed dense 3-D reconstruction method excelled in producing accurate 3-D point clouds of various plants while retaining colors, textures, shapes as compared to other current methods, i.e. PMVS, CMPMVS, 3DSOM. It is flexible and efficient. Abundant useful phenotypic data as plant height, plant topology, stem width and length, numbers of leaves, leaf area and leaf angle, etc. can be extracted.

(a) (b)

(c)

(d)

Figure 5.12: 3-D reconstruction of an *Arabidopsis*. (a) raw image (one of 66 images); (b)(c)(d) 3-D reconstructions by CMPMVS, PMVS and the proposed method, respectively.

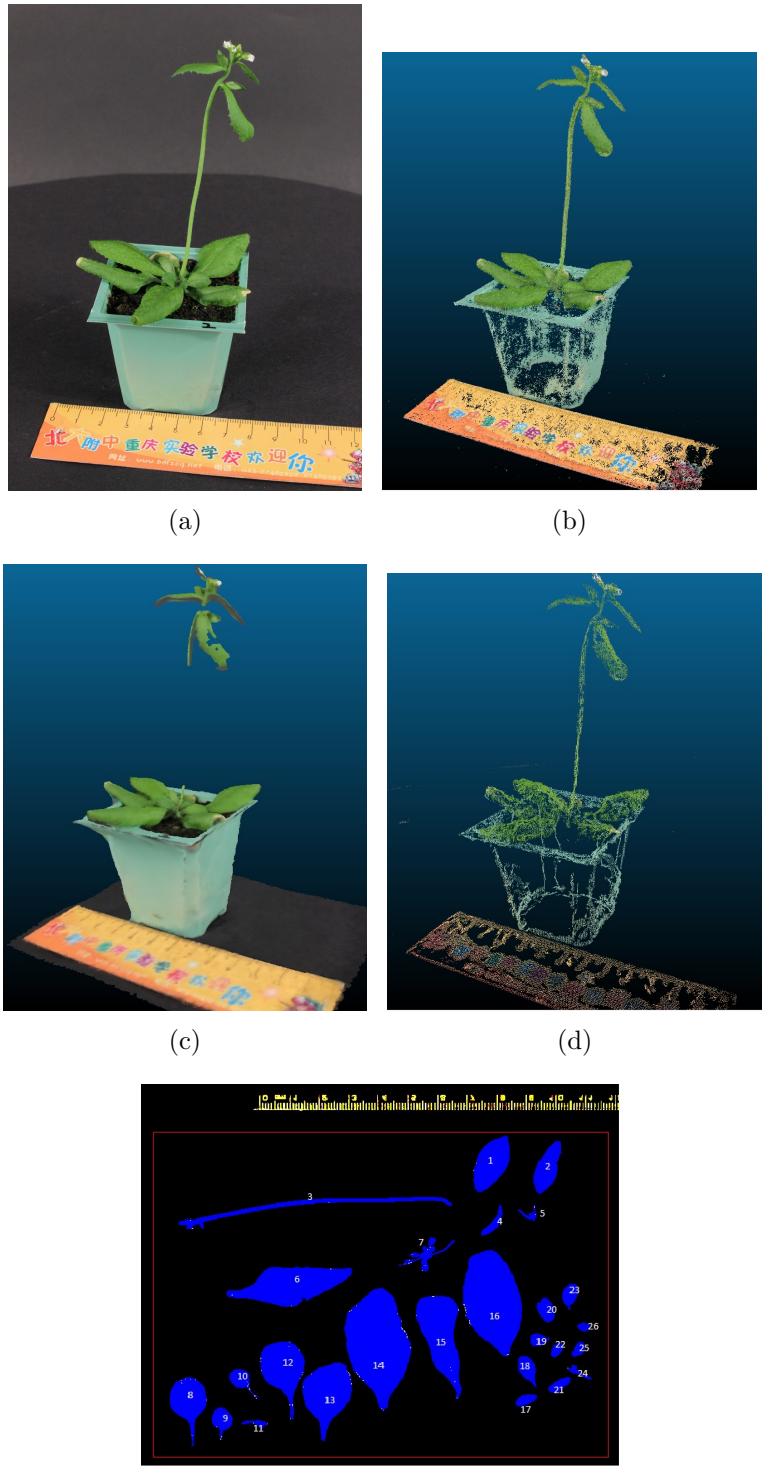



Figure 5.13: (a) Raw image of an *Arabidopsis* (one of 62 images); (b)(c)(d) 3-D reconstruction by the proposed method, CMPMVS, PMVS, respectively; (e) illustration of the measurement of the leaves and stems using *WinDIAS* system.

# Chapter 6

# Automatic Phenotypic Feature Extraction from 3-D Point Cloud of Plants

The "phenotype" is composed of an enormous amount of processes, functions and structures, which change during growth and development. The plant body rapidly becomes more complex due to re-iterative organ formation, changes in organ spacing and identity, and branching that leads to overlapping and variable 3-D organization. This chapter aims to develop cost-effective 3-D modeling of plant growth. I proposes an unsupervised 3-D point cloud segmentation approach, as well as a measurement method of branch angle for the understanding and analysis of the structure of plants of interest.

## 6.1 Introduction

Functional-structural plant models (FSPMs) seek to integrate geometric structure with function ( such as the flow of material, energy or information through a system) to understand how this is affected by the genotype and as modified by environmental influences. FSPMs attempt to simulate growth and development through modelling the development, growth and function of individual cells, tissues, organs and plants in their spatial and temporal contexts. Accurate models are often associated with 3-D visualization of the plant architecture

and FSPMs (or virtual plants) attempt to explicitly describe development over time (Prusinkiewicz & Hanan, 1990) (Sievänen *et al.*, 2000) (Vos *et al.*, 2007) (Vos *et al.*, 2010). Early the most influential work, Lindenmayer-Systems (commonly named *L*-systems), was done by Lindenmayer (Lindenmayer, 1968) and Prusinkiewicz et al. (Prusinkiewicz & Hanan, 1990). *L*-systems are a formal language developed to describe both static plant structure and its dynamics (driven by functions) in the form of a set of explicit rules. The central idea of *L*-systems consists of rewriting of a string of modules representing the structure of the plant. Explicit rewriting rules express the creation and changes of state of the modules over time. Such rules can be expressed using a dedicated programming language ((Prusinkiewicz *et al.*, 1999)) or by incorporating *L*-system based language constructs into existing languages, such as C++ (L+C) (Karwowski & Prusinkiewicz, 2003), Java (GroIMP)(Kniemeyer & Kurth, 2007)) or Python (L-Py, PlantGL and GreenLab)(Pradal *et al.*, 2008; Pradal *et al.*, 2009; Boudon *et al.*, 2012).

According to the formalism of *L*-systems, a plant is viewed as a developing assembly of individual units, or modules. These modules are characterized by parameters such as length, width, and age, as well as parameters characterizing shape. A methodology for constructing *L*-system models based on empirical estimates of such parameters has been introduced by (Prusinkiewicz *et al.*, 1994).

Many papers describe the use of different types of equipment (e.g. rulers, protractors, sonic or magnetic digitizers) to collect data on the spatial orientation and shape of organs and how to process such data to arrive at mathematical descriptions of shapes and angles and, finally, at the reconstruction of a simulated plant structure in silico (Sinoquet *et al.*, 1997) (Drouet, 2003) (Evers *et al.*, 2005) (Dornbusch *et al.*, 2007) (Dauzat *et al.*, 2008) (Zheng *et al.*, 2008).

An empirical model of Arabidopsis was presented in (Mündermann *et al.*, 2005), which simulates and realistically visualizes development of aerial parts of the plant from seedling to maturity. This model integrates thousands of measurements data including sizes and shapes of individual organs (internodes, leaves, and flower organs,etc.), taken from several plants at frequent time intervals destructively and non-destructively. A comparison of selected devel-

opmental stages of an individual plant over time was made in (Mündermann *et al.*, 2005) (Fig. 6.1). The comparison indicates that the model captures the architecture of a growing Arabidopsis plant faithfully, but there is obviously difference between 3-D model and the real plant, because that the values of measured parameters are averaged over several plants, as well as some empirical parameters in *L*-systems lack explicit physical meanings (Mündermann *et al.*, 2005).

Four contrasting alfalfa plants were manually measured using an electromagnetic 3-D digitizer (3Space Fastrak, Polhemus Inc., Colchester, VT, USA) in a 3-D digitizing experiment (Gaëtan *et al.*, 2012), where the detailed records, including the multi-scale topology of each plant, the spatial coordinates and geometry of all plant organs, were performed as described in (Sinoquet & Rivet, 1997; Barillot *et al.*, 2011), and then virtual plants were reconstructed by triangulation processing using the PlantGL toolkit (Pradal *et al.*, 2008). In (Gaëtan *et al.*, 2012), for each of the digitised plants, its 3-D point clouds were acquired using a rapid non-contact 3-D laser scanner (VIVID 910, Konica Minolta Holdings Inc., Tokyo, Japan). Plants were placed on a synchronised rotary table and captured every 120°, and the acquisition data were merged without outlier processing using the "Rapid Form"software (Inus Technology Inc., Seoul, South Korea), the resulting 3-D point clouds were further processed with MeshLab (MeshLab, 2014). Fig. 6.2 shows the experimental results. The procedure of 3-D reconstruction and analysis, involving manually data collection, laser scanning and 3-D data merging, and post-precessing of the 3-D point clouds, were expensive and time-consuming.

The ability to capture efficiently accurate 3-D information from a growing plant at any given time is attractive for both FSPM and plant phenotyping. Such models can contain the information needed to compute a variety of plant traits, such as total leaf shape and area, branch angle, etc., and are essential where other data (fluorescence, thermal etc) needs to be co-registered with the complex surface morphology of the plant. Therefore, there is an urgent need for rapid, automated, and generalized techniques for accurately reconstructing and measuring the 3-D architecture of complex plants.

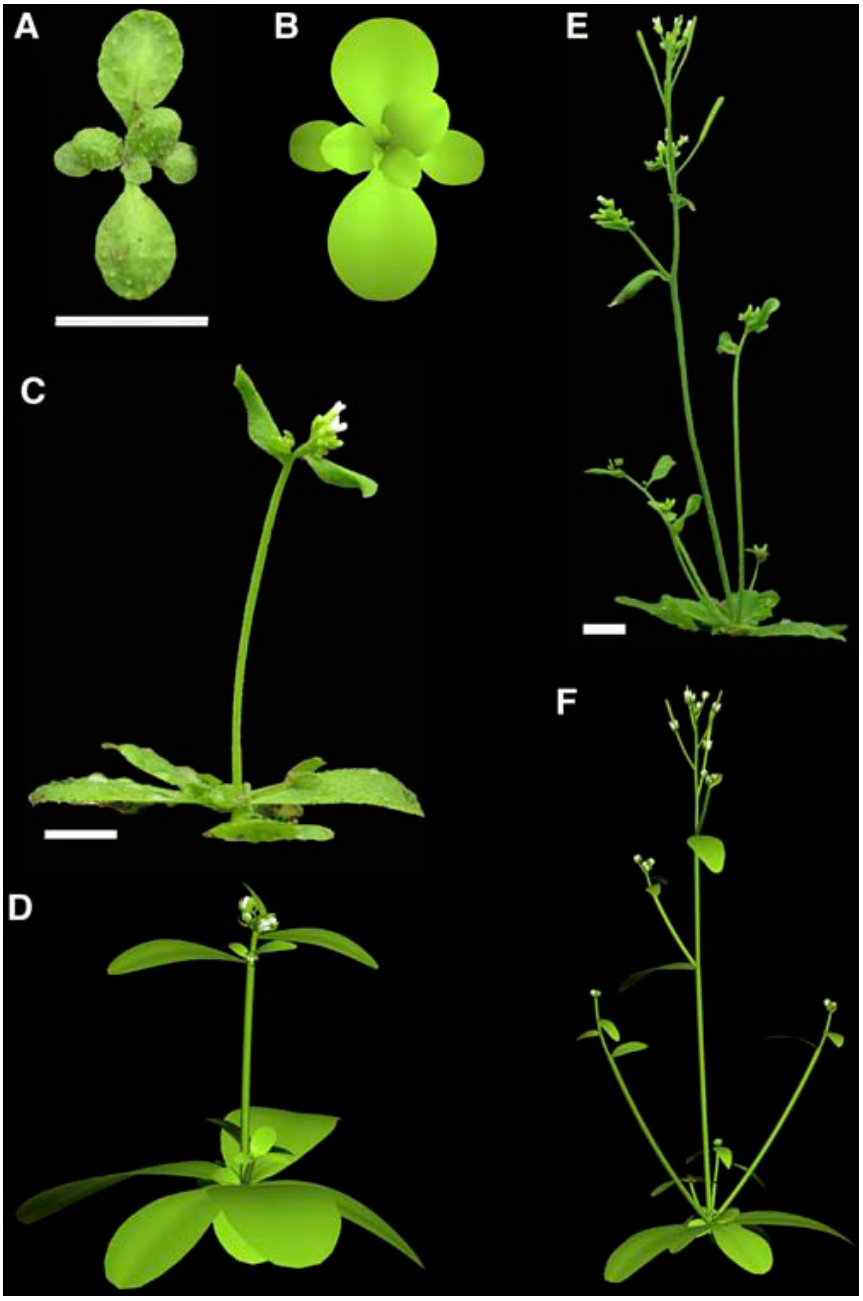



Figure 6.1: Comparison of sample Arabidopsis plants (A, C, E) with the model (B, D, F) after different hours from seeding (HFS). A and B, at 264 HFS; C and D, at 417 HFS; E and F, at 491 HFS. Scale bar =1cm. The figure is taken from (Mündermann *et al.*, 2005).

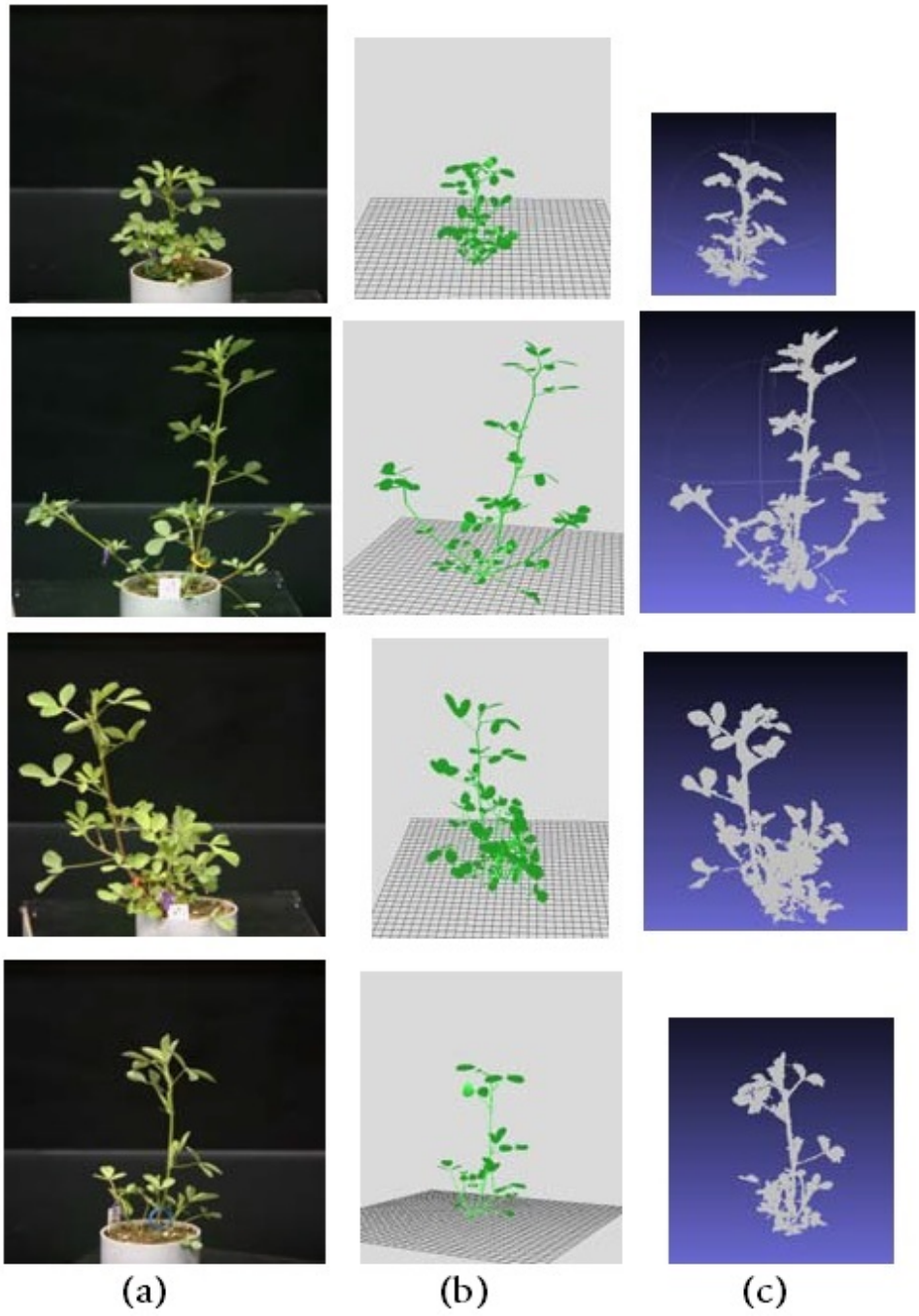



Figure 6.2: 3-D digitizing experiment of four alfalfa plants. (a) images; (b) corresponding 3-D digitized models; (c) 3-D point clouds produced by 3-D laser scanner. The images are taken from (Gaëtan *et al.*, 2012).

## 6.2 The Proposed Method

My goal aims at developing suitable measurement techniques that can be applied to diverse plant species, of variable architectural complexity. The pro-

posed methods are described in the following sections.

### 6.2.1 3-D Data Acquisition of Plants

I used a single camera (Canon 600D with *18-55mm* variable focus lens) and turn-table to capture tens of thousands of plant images ($3696 \times 2448$ high-resolution) without calibration required in advance, and built up a dataset of 3-D plant point cloud of various plants (including *Arabidopsis thaliana*, Oat, Wheat, Maize, Foragegrass (*Lolium*), Clover, Physalis, and Brassica).

### 6.2.2 Pre-processing for the Raw 3-D points

After 3-D acquisition of plants, I obtained the dense 3-D point clouds that include up to millions points with associated color information. However, there are large numbers of redundant points and a small quantity of random noise (outlier points) in the dense 3-D point cloud. Pre-processing, such as noise removal, down-sampling, etc., are necessary before further analysis.

*Noise Removal*: Because I captured the multiview images of plants in controlled environment, the colors of most of 3-D noisy points in the raw dataset are close to black (we used a black curtain as background) and can be clearly distinguished from those 3-D organ points whose colors are almost green. Therefore, I adopted a color filter to remove the 3-D noisy points efficiently.

*Down-sampling*: In my work, while such a large number of points are useful to describe the details of a plant of interest, it is computationally expensive to process and analyze such data (up to millions 3-D points). To reduce the computational complexity, a sampling method was used to down-sample the raw 3-D point cloud to a sparse point cloud with less than twenty thousand points. Considering that the 3-D point cloud is incomplete (caused by occlusion) and densities are not even, I employed an adaptive down-sampling method, instead of uniform down-sampling, where the neighboring points were merged together if those points are nearer than the specified distance threshold. The threshold used in the proposed work is not fixed value, depending on different application, it can be adjusted from 0.005% to 0.05% of diagonal distance of the enclosing box that surrounds the 3-D points.

## 6.3 Segmentation of 3-D Points of Plants

The problem of segmenting and skeletonizing plant structure in 3-D point clouds remains challenging due to many reasons: occlusion, imaging noise, thin and pseudo-linear structure, and closeness of one part to another. Previous work usually focused on scenes (outdoor/indoor), buildings, man-made objects like toys, desks, chairs, cars, or CAD models. Most of these subject have abundant characteristics, such as flat surfaces, colors, textures, well-defined geometric sizes or shapes (plane, cylinder, sphere, etc.), which can be helpful for segmentation of the single, usually discrete, object. Many algorithms have been proposed to decompose an object into functionally meaningful parts or regions, including region growing, $K$-means, hierarchical clustering and spectral clustering (Liu & Zhang, 2004) (Shlafman *et al.*, 2002) (Shamir, 2006). See Fig. 6.3.

For segmentation of 3-D point clouds of plant material, most work focused on detecting different organs using features training and classfication, or by means of known information like leaf size, stem size, colors, etc., and so these methods were only applicable to limited plants (Paulus *et al.*, 2013) (Dey *et al.*, 2012) (Paulus *et al.*, 2014a).

In recent years, spectral clustering has become one of the most popular modern clustering algorithms, which treats the data clustering as a graph partitioning problem without make any assumption on the form of the data clusters. Spectral clustering examines the connectedness of the data, whereas other clustering algorithms such as $K$-means use the compactness to assign clusters, and makes use of the spectrum (or eigenvalues) of the similarity matrix of the data (Von Luxburg, 2007).

Considering that the resulting 3-D point clouds of plants are dense, complete and have good local connectedness, I used spectral clustering to segment the sampled points.

*Spectral Clustering* (Von Luxburg, 2007) is described by an undirected complete similarity graph $G = (V, E)$ where $V$ is a set of vertices $v_i$ with one vertex for each of $n$ points in the original data and $E$ is an edge set of similarities $s_{ij} = s_{ji}$ between pairs of points. The $s_{ij}$ is defined by user to denote local neighborhood relations between data points, such as distance, etc. If $s_{ij}$ is larg-

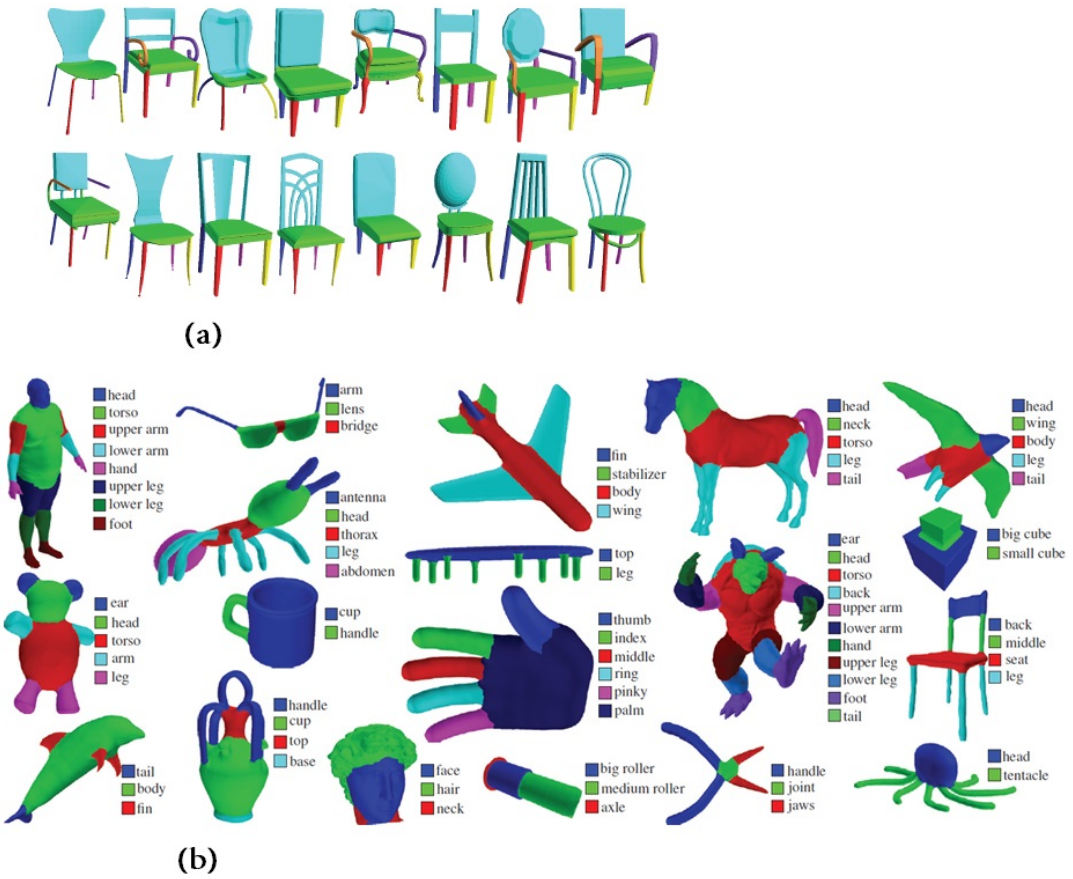



Figure 6.3: Illustrations of 3-D segmentation of objects. (a) segmentations of a set of chairs; (b) 3-D mesh segmentations for every object category in the Princeton Segmentation Benchmark (Chen *et al.*, 2009) .The images are taken from (Golovinskiy & Funkhouser, 2009; Kalogerakis *et al.*, 2010) respectively.

er than a certain threshold, two vertices are connected and the edge is weighted by $s_{ij}$. The weighted adjacency matrix $\mathbf{W}$ is the $n \times n$ symmetric matrix of pair-wise similarities between points as defined:

$$\mathbf{W} := (w_{ij})_{i,j=1,\ldots,n} \geq 0 \tag{6.1}$$

The degree of a vertex $v_i \in \mathbf{V}$ is defined as

$$d_i := \sum_{j=1}^{n} w_{ij} \tag{6.2}$$

Degree matrix $\mathbf{D}$ as the diagonal matrix with $d_1, ...d_n$ on the diagonal.

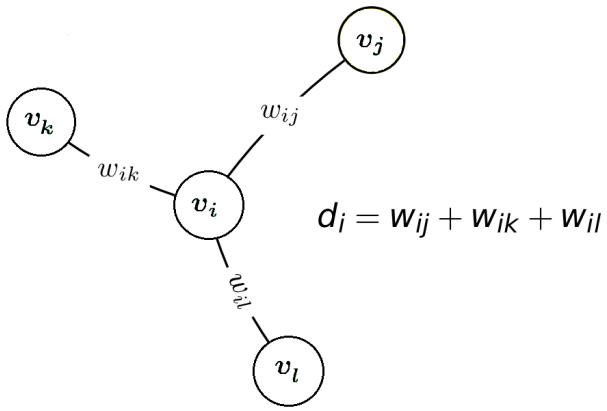



Figure 6.4: Similarities and degree of graph.

The degree matrix $\mathbf{D}$ summarizes the similarity of each vertex with respect to the other vertices as a whole, whereas the similarity matrix $\mathbf{W}$ summarizes pair-wise similarities between vertices. To compute spectral features, $G$ is expressed as an $n \times n$ Laplacian matrix. There are a number of definitions for the Laplacian matrix (Von Luxburg, 2007), but here, I used the simplest one, $\mathbf{L} = \mathbf{D} - \mathbf{W}$, where the matrix $\mathbf{L}$ is symmetric, positive semi-definite with smallest eigenvalue equal to 0. To construct spectral features, the eigenvectors associated with the next $k$ smallest eigenvalues of $\mathbf{L}$, are formed into an $n \times k$ matrix $\mathbf{F}$. Spectral features associated with the $i^{\text{th}}$ vertex, are taken from the $i^{\text{th}}$ row of $\mathbf{F}$.

Generally speaking, all the spectral clustering algorithms applied in (Shamir, 2006) (Liu & Zhang, 2004) are intrinsically similar, and the main differences are in creating the similarity matrix and in how to select the parameters such as thresholds and number of clusters. My method differs from previous work (Liu & Zhang, 2004) which focus on 3-D mesh segmentation for CAD models:

(1) Considering shape complexity, sufficient 3-D points are needed to represent a plant in my dataset where there are $10k \sim 50k$ down-sampling points, much more than the usual few thousand points in typical 3-D CAD models.

(2) I created a graph using K-Nearest Neighbors ($K$-$NN$) and thus each point will be connected to all its $K$ neighbors (here I used $K = 20$).

(3) I constructed a similarity matrix using $K$-$NN$ distance and normal smoothness and then employed spectral clustering to cluster points into initial segments. In its $K$-means clustering step, I intentionally selected $K = 75$ ($K$ can be chosen from 50 to 100) to get initial clusters with over-segmentation.

(4) I combined neighboring segments according to their normal similarity (threshold 10°) to produce the final segmentation results.

The proposed method is based on Spectral Clustering (see (Von Luxburg, 2007) for details), but the parameters ($K$ and threshold) were chosen empirically in the experiments. Although the same parameter values were used in the experiments for six different plants and the final segmentation results were also insensitive to the minor adjustments of the parameters, since plants have diverse structures which change during growth and development, how to optimize the parameters will obviously be a key problem, especially for full-automatic segmentation on 3-D points of plants, in the future work aiming at high-throughput plant phenotyping.

The proposed method is described in Algorithm. 1.

## 6.4 Estimation Angle of Branches

The angle between two branches is defined as the angle included by the two axes of the said branches, where the branches can be considered to approximate to cylinders. Branch angles are very important parameters that directly determine the structure of a plant in an $L$-system-based model. Usually, the angle was supposed as initial constant empirical value with a fixed changing rate but the angle of an actual plant can be perturbed by the act of measurement. The branch angle of a real plant is also affected by environmental factors. Therefore, to measure the angle of branch directly from 3-D point cloud is very attractive.

### 6.4.1 Skeleton of 3-D Point Cloud of Plant

I firstly made an effort to compute the branches' angle through abstracting the skeleton of 3-D plant points. Unlike some man-made objects, where the

**Algorithm 1** The segmentation of 3-D point cloud.

**Input:** point cloud $V = \{v_i \in \mathbb{R}^3, 1 \leq i \leq n\}$.

1: to construct similarity matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ using K-NN distance and normal smoothness as similarities, let $\mathbf{W}$ be its weighted adjacency matrix and $\mathbf{D}$ the degree matrix.

2: compute the Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{W}$.

3: compute the first $k$ eigenvectors $u_1, \ldots, u_k$ of $\mathbf{L}$.

4: let $\mathbf{U} \in \mathbb{R}^{n \times k}$ be the matrix containing the vectors $u_1, \ldots, u_k$ as columns.

5: for $i = 1, \ldots, n$, let $y_i \in \mathbb{R}^k$ be the vector corresponding to the $i$-th row of $\mathbf{U}$.

6: cluster the points $(y_i)_{i=1,\ldots,n} \in \mathbb{R}^k$ with the $K$-means algorithm into clusters $C_1, \ldots, C_k$.

7: clusters $A_1 \ldots A_k$ with $A_i = \{j_i | y_i \in C_i\}$.

**Output:** combined neighboring clusters to produce the final segmentation results according their normal similarity.

skeletons of objects are sometimes ambiguous or even meaningless, the skeleton of a 3-D plant point cloud can clearly represent the structure of plant. But extraction of the skeleton of a leaf is a challenging problem because of its variable contour and intrinsic thinness. However, for $L$-systems, it is only necessary to define the contour of a leaf and not its skeleton. Therefore, when leaves are segmented, I manually remove them from point cloud and then adopt a skeletonisation process to extract the basic branch-structure of the plant. Due to occlusion, the plant point cloud is normally incomplete, and the density of points varies from one region to another. In order to extract a skeleton with complete connectivity and reasonable accuracy, the topological thinning method proposed by Cao *et al.*(Cao *et al.*, 2010) was adopted. This method is based on Laplacian contraction (see (Cao *et al.*, 2010) for details), and it has the clear advantage that not only is it robust to noise but it also produces a well connected skeleton.

However, I found that the skeletons of two neighboring branches can not always accurately represent the medial axis (a case shown in Fig. 6.9) so that it was difficult to calculate the expected angle.

### 6.4.2 Estimation of Branches Angle using Bounding Boxes

3-D Bounding boxes are three-dimensional boxes (eight corners) that surround entities within the measuring model. The mini-3D Bounding box is such box, which has a minimum volume, and was proven that at least its two adjacent box sides are flush with edges of the hull (O'Rourke, 1985). For simplicity, a 2-D Bounding box is illustrated in Fig. 6.5, where a minimal bounding rectangle can be found by means of rotating the convex hull and calculating the minimum rectangle area.

It is a basic assumption in the proposed segmentation method that the 3-D mini-bounding box surrounding a branch (cylinder) has a longest edge that parallels the axis of branch. Therefore, I firstly employed 3-D mini-bounding box algorithm to obtain the mini 3-D bounding box corresponding to each segmented "branch" (chosen manually because organ classification has not been implemented in this project), and then I calculated the intersection angle of two longest edges of the two mini-bounding boxes (as shown in Fig. 6.10). I supposed this intersection angle could approximate to the angle of two branches (see Fig. 6.11). The method is described in Algorithm.2.

**Algorithm 2** Branch angle estimation based on mini-boundingbox

**Input:** Segmented point cloud $V$.

1: Manually identify those segments of interest that can represent the branches.
2: **for** each face of the convex hull **do**
3: Compute the face's orientation and rotation matrix.
4: Rotate the convex hull to x-y plane.
5: Find out min/max values of x/y/z of the rotated convex hull and compute the volume of the enclosing box.
6: **end for**
7: Find out two 3-D bounding boxes with the minimum volume
8: Find out the longest edges of two mini 3-D bounding boxes

**Output:** The intersection angle of the longest edges.

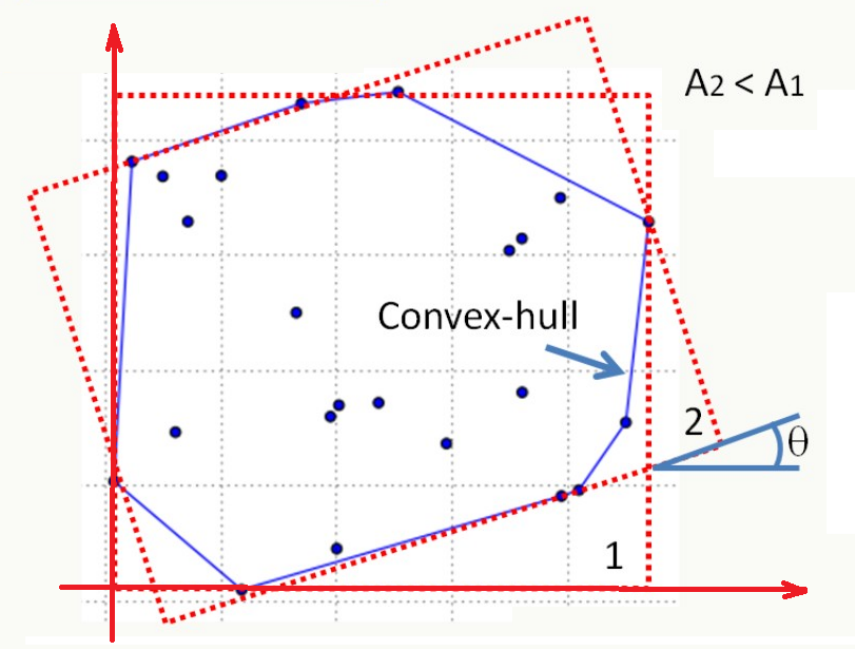


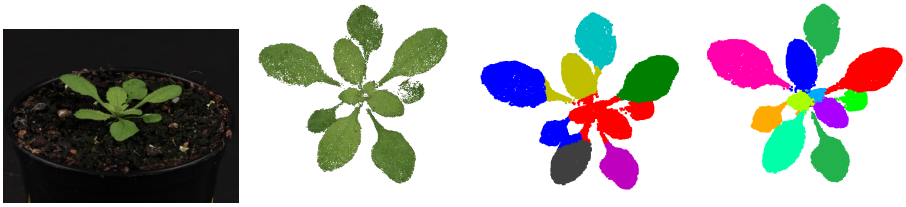

Figure 6.5: 2-D mini-Bounding Box

## 6.5 Experimental Results and Analysis

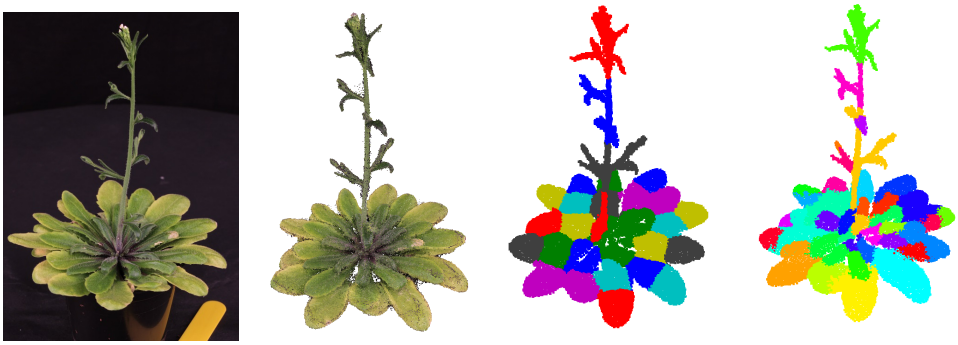
The experiments were performed on an Intel Core i7 2.7GHz machine with 16 GB RAM. In the experiment of spectral clustering segmentation, the majority of the computational time is actually spent on constructing the similarity matrix, which depends on the number of 3-D points, but is typically around 6 minutes for $20K$ points. Comparatively, the clustering time is almost negligible. The segmentation results of selected plants are shown in Fig. 6.7.

I had also compared the proposed segmentation method with the standard $K$-means algorithm. From comparison results in Fig. 6.7, it can be seen that obviously the $K$-means just cluster the points in proportional spacing without considering the local connectivities, and the proposed method excelled the $K$-means in each selected plant. Furthermore, the $K$-means algorithm requires a manual input as to the number of cluster $K$, but this is certainly a difficult issue especially for complex structure like plants. In the experiments, I just empirically chose the $K$ from 10 to 30 depending to the structural complexities of the selected plants.

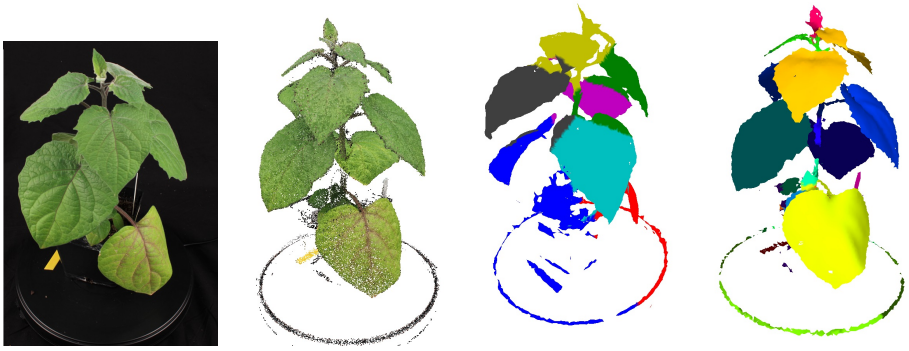
Quantitative evaluation of segmentation quality is difficult because either



(a) (left to right): Arabidopsis rosette (3 weeks growing, 1 of 60 images); 148857 3-D points; result of $K$-means (chose $K$=10); result of the proposed method.



(b) (left to right): flowering Arabidopsis (6 weeks growing, 1 of 60 images); 712540 3-D points; result of $K$-means (chose $K$=30); result of the proposed method.



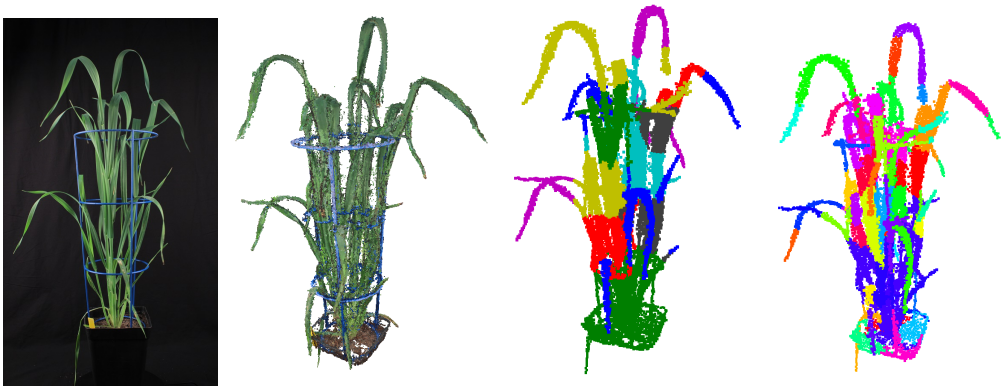
(c) (left to right): Physalis plant (1 of 90 images); 322020 3-D points; result of $K$-means (chose $K$=10); result of the proposed method.

Figure 6.6: The 3-D segmentation results of various plants. 1st column: one sample of the 2-D plant images; 2nd column: 3-D points after down-sampling; 3rd column: segmentation result by standard $K$-means; 4th column: segmentation result by the proposed method.

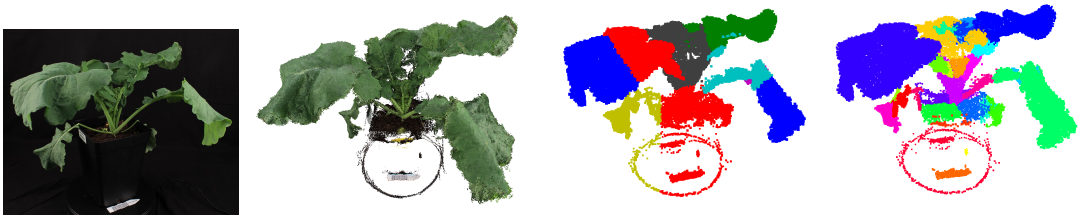
the lacking of the ground truth, or the effectiveness of the segmentation is determined by the application for which it is intended. However, I found that:

(1) The proposed approach is applicable to diverse plant species − only use raw 3-D point data as the input not considering the other information such as shape, color, size, etc.;

(a) (left to right): maize (1 of 120 images) ; 906012 3-D points; result of $K$-means (chose $K$=20); result of the proposed method.



(b) (left to right): wheat (1 of 120 images) ; 2128586 3-D points; result of $K$-means (chose $K$=20); result of the proposed method.



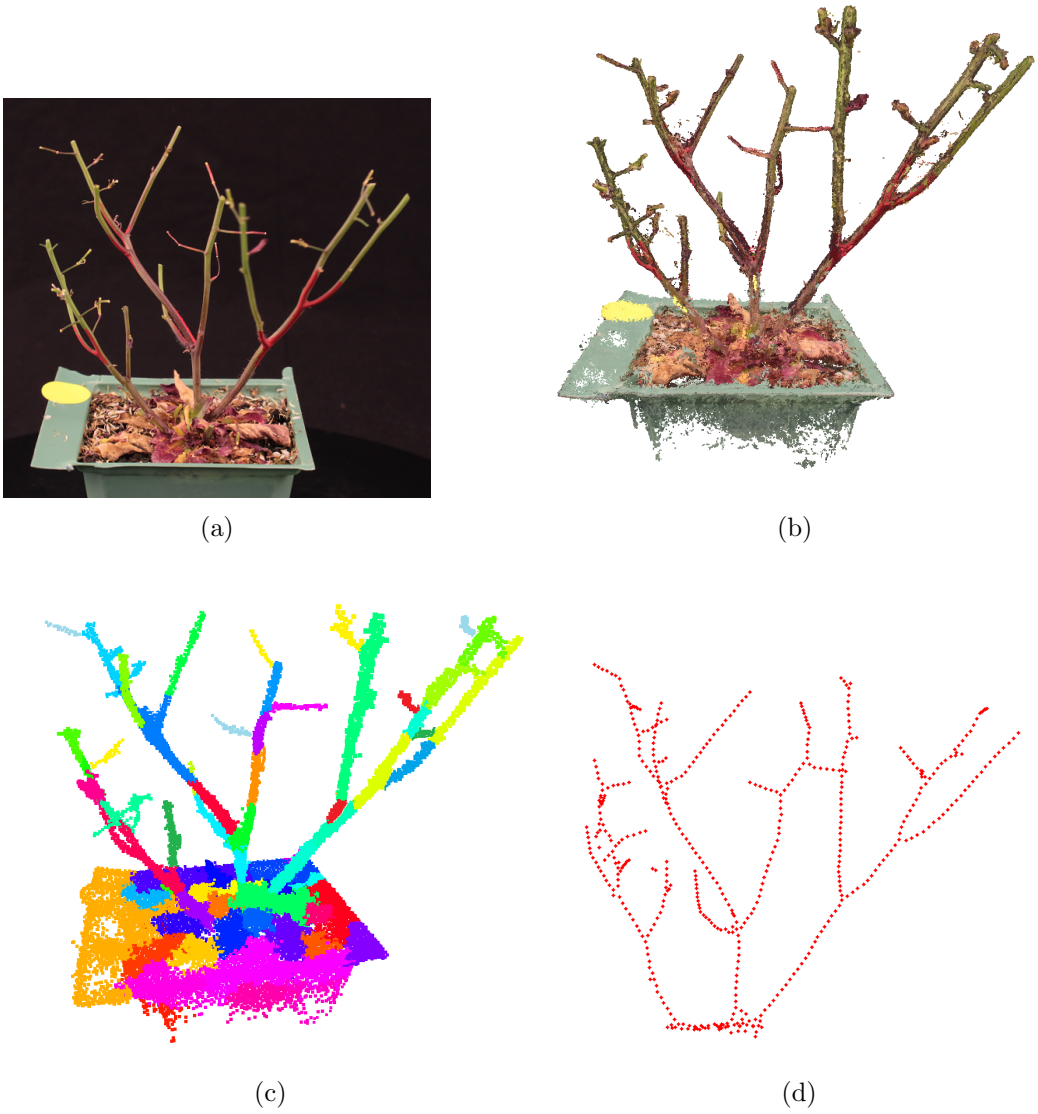
(c) (left to right): brassica (1 of 120 images) ; 1588773 3-D points; result of $K$-means (chose $K$=10); result of the proposed method.

Figure 6.7: The 3-D segmentation results of various plants. 1st column: one sample of 2-D plant images; 2nd column: 3-D points after down-sampling; 3rd column: segmentation results by standard $K$-means; 4th column: segmentation results by the proposed method.

(2) Satisfying segmentation on 3-D points of plants with various shapes is really a rather difficult problem – it is not always possible to get meaningful and accurate segmentation, especially for leaves with curved surfaces, or tiny side-branches at the top of an *Arabidopsis* flowering stem, and at most junctions points of a plant.

The skeletonizatoin result of a sample plant is illustrated in Fig. 6.8(c). It

can be seen that the skeletons produced are representative of the plant structure. Although the skeletons cannot used to estimate the angle of branches in this work, the plant skeleton cloud be potentially parsed to syntax expression to define the architecture rules of $L$-systems in further researching.



(a) (b)

(c) (d)

Figure 6.8: (a) sample of 2-D images of an *Arabidopsis* (leaves cut off) (b) 3-D point cloud. (c) segmented 3-D points. (d) 3-D skeleton points.

I used the plant model in Fig. 6.8(b) to obtain a set of measured angles, then compared those with the values obtained by manual measurement and provided by a biologist. The measurement results of a number of branch angles are shown in Table. 6.1, and the root mean square error (RMSE) of the estimated angles

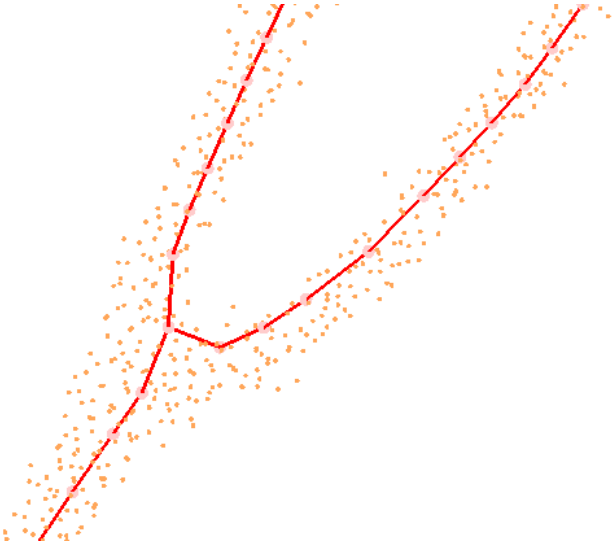


Figure 6.9: The skeleton of two branches.

is 1.88°.

To the best of my knowledge, there does not exist any methods in the literature for angle estimation from 3-D plant points. Thus I used a protractor as a measuring reference and produced a set of its 3-D point clouds with different angles varied from 20° to 90°, and then calculated the angles of its bars using the proposed method (shown in Fig. 6.12). The results are shown in Table. 6.2, and the root mean square error (RMSE) of the estimated angles is 0.45°. These results show that the proposed method is reliable. Nevertheless, it should be noted that this method is dependent on the segmented branches, since segmented branches are identified and chosen manually.

| Item | Ground truth | Measured value | Relative error(%) |
|---|---|---|---|
| angle1 | 59.80 | 58.70 | -1.89% |
| angle2 | 52.40 | 49.40 | -5.73% |
| angle3 | 61.90 | 63.30 | 2.26% |
| angle4 | 69.20 | 68.57 | -0.91% |
| angle5 | 39.80 | 42.06 | 5.68% |
| Error (RMSE) | | 1.88 | |

Table 6.1: RMSE of estimated angles (in degrees).

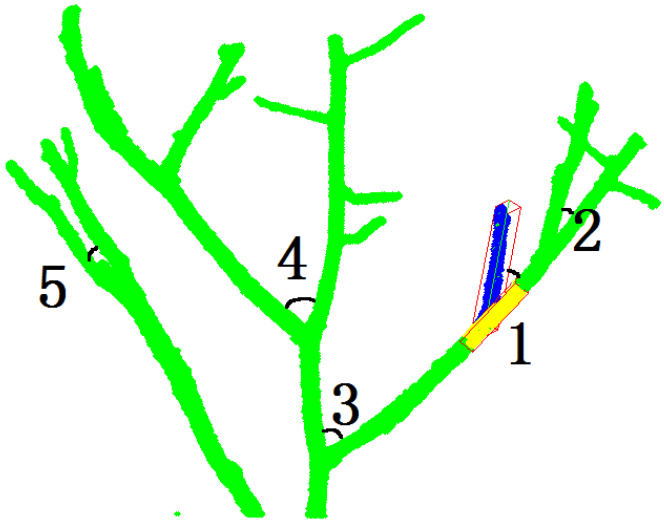



Figure 6.10: The angle included by mini-bounding boxes.

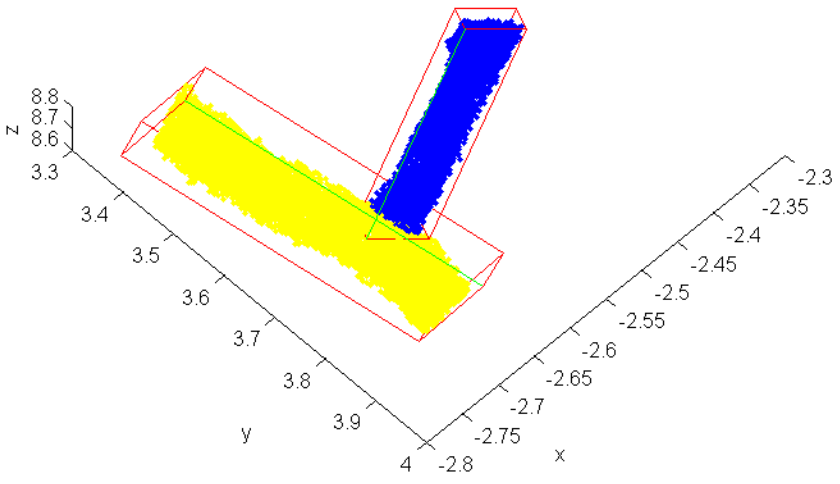
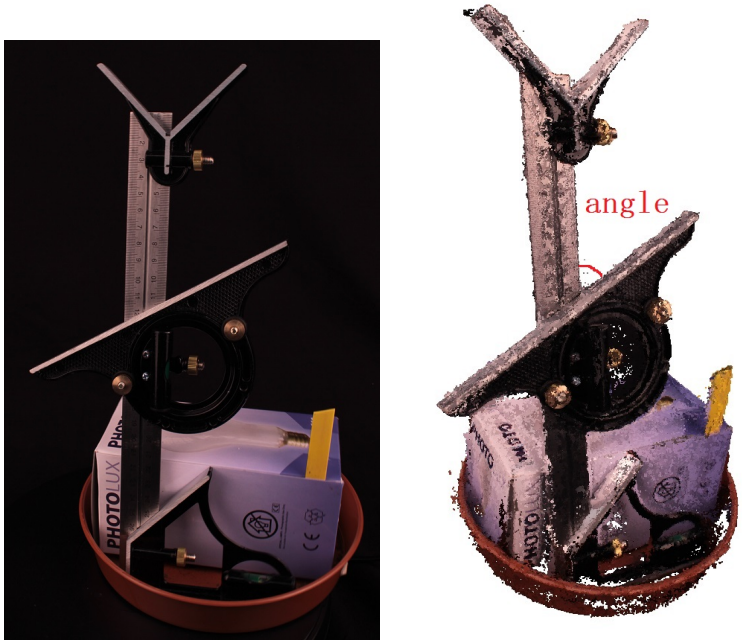



Figure 6.11: Two mini-bounding boxes approximate two branches.

For the proposed segmentation method, the main limitation is that component recognition could be further refined − for example, the leaves sometimes contain aspects of petiole or stem, depending on plant morphology and architectural complexity. A fully automated 3-D segmentation method that can cope with a wide range of different shaped plants is still a rather challenging




(a) (b)

Figure 6.12: (a) Raw 2-D image of an protractor (b) 3-D point cloud of the protractor.

problem, because segmentation is often determined by points and a meaningful segmentation on a plant can be quite different in other plant species or morphological mutants. An efficient segmentation method may require a combination of machine learning, prior knowledge or user intervention.

## 6.6 Summary

This chapter has proposed a novel automatic 3-D point cloud segmentation method through adapting the spectral clustering algorithm, which is applicable to diverse plants with varied structure, size and shape, and then has presented an interactive approach on estimation of angles between branches obtained from the previous segmentation step. The angle estimation method is relatively easier compared with analyzing the parameters of other organs. Future work could focus on developing automatic methods to measure various organs of plants, improving analysis ability of plant phenotyping, and bridging actually measured values with the parameters in FSPMs to ensure close relationships

| Item | Ground truth | Measured value | Relative error(%) |
|---|---|---|---|
| angle-20° | 20 | 19.47 | -2.64% |
| angle-22° | 22 | 22.33 | 1.52% |
| angle-30° | 30 | 30.45 | 1.50% |
| angle-38° | 38 | 38.34 | 0.88% |
| angle-45° | 45 | 44.33 | -1.48% |
| angle-50° | 50 | 50.42 | 0.85% |
| angle-60° | 60 | 60.24 | 0.39% |
| angle-70° | 70 | 70.56 | 0.79% |
| angle-90° | 90 | 89.65 | -0.38% |
| Error (RMSE) | | 0.45 | |

Table 6.2: RMSE of estimated angles (in degrees)

between FSPMs and real plants.

# Chapter 7

# Conclusions

This chapter summarizes and concludes the investigations conducted in this thesis, including a list of the main contributions together with a list of publications produced from this project. At the end of this chapter, a brief discussion about possible future directions of research is also presented.

## 7.1 Summary of the Thesis

Phenotyping aims to measure complex traits related to growth, yield and adaptation to stress with a certain accuracy and precision at different scales of organization, from organs to canopies, and potentially across time. Traditionally, this is either limited to tedious and sparse manual measurements, often acquired destructively, or coarse image-based 2D measurements. In recent years, high throughput phenotyping platforms have been deployed in growth chambers or greenhouses. These platforms use robotics, precise environmental control and imaging technologies (hardware and software) to assess plant growth and performance. However, these platforms are designed for a limited range of species, encompassing small rosette plants such as *Arabidopsis* and the primary cereal crops. Generic platforms and solutions enabling the simultaneous phenotypic evaluation of multiple species must be developed.

In this thesis, I explored the existing 2-D/3-D techniques in computer vision and have developed a low-cost multi-view image based 3-D reconstruction system for non-destructive plant phenotyping.

In the system, according to the requiremants, existing resources and limitations of most biology labs, the data acquisition was designed as a manual procedure (within 1∼2 minutes) of capturing image with low-cost turntable and camera that aims to address the occlusion problem of plants, where the quality and flexibility of multi-view 2-D image acquisition are more important than the degree of automation. In order to estimate the camera's parameters and poses from uncalibrated multi-view 2-D images, a full-automatic SFM was applied and its computation speed was important (within 10∼30 minutes mainly depending on the number of images). Once the SFM outputs the results successfully, a full-automatic 3-D reconstruction was utilized to produce dense and detailed 3-D point clouds of plants and its speed was less crucial (within 15∼90 minutes mainly depending on the number of eligible stereo image pairs). For the performance of the whole system, the accuracy of measurement of a specific phenotypic parameters from the final 3-D point clouds was most important to biologists, instead of the accuracy of the SFM or the speed of the 3-D reconstruction.

The whole work can be divided into four phases: development of multi-view 2-D image acquisition system (Chapter 3), camera parameters estimation from uncalibrated multi-view 2-D images (Chapter 4), accurate multi-view stereo 3-D reconstruction (Chapter 5) , and an approach to extract the meaningful geometric features of organs (Chapter 6), which consists of an automated segmentation method and a semi-automated measurement of branch angles from resulting 3-D point clouds of plants.

I started with a brief introduction of the research topics under investigation in this thesis, covering the background of plant breeding/plant phenotyping, the overview of limitations of existing 2-D/3-D imaging techniques applied in plant phenotyping platforms, and the challenges caused by plants' complex shapes, self-changing structures and self-occlusions. A review of the relevant literature and background was carried out in the following chapter, including existing 2-D/3-D imaging techniques applied in plant phenotyping systems, the technical characteristics of current 3-D sensors and some related topics of digital camera, camera model and multi-view geometry.

In order to build a complete framework for 3-D reconstruction using low-

cost equipments and resources (digital camera, turn-table and free software library), I first designed a practicable multi-view 2-D image acquisition system to resolve the occlusion problem as much as possible, which consists of some common digital photography equipment, including DLSR camera, tripod, lighting, background curtain and computer, and is capable of capturing high-resolution images of a variety of plants under their different growth stages.

Subsequently, I developed a camera parameters and pose estimation method from the captured multi-view images of plants, which includes the following steps: i) extraction of local invariant features for each image; ii) feature matching from neighboring overlapped image pairs, and iii) estimation of camera parameters and pose based on incremental Structure from Motion.

For dense 3-D reconstruction of plants, I developed a multi-view stereo based 3-D reconstruction method, which consists of stereo pair selection, depth maps computation, depth maps refinement and depth maps merging, and is able to yield more dense, complete and detailed 3-D point clouds than the existing state of art dense 3-D reconstruction methods, such as PMVS and CMPVS, that have been extensively used in many 3-D reconstruction objects (e.g. city scene, building, statue, heritage). To evaluate the quality of resulting 3-D point cloud, I extracted interactively some phenotypic information of organs (e.g. stem length, leaf area) from the 3-D point cloud using 3-D software, and compared them with ground truth as measured by traditional destructive harvest. The comparison analysis results show that the accuracy of the proposed method is high enough to distinguish different genotypes plant and precision is stable enough to adapt slightly deviated experimental conditions. I concluded that the proposed 3-D reconstruction is a promising generalized technique for the non-destructive phenotyping of various plants during their whole growth stages.

Finally, I proposed a novel automatic 3-D point cloud segmentation approach through adapting the spectral clustering algorithm, which is applicable to diverse plants with varied structure, size and shape, and then has presented an interactive approach on estimation of angles between branches obtained from above segmentation step. Experimental results show that the approach is efficient and reliable, and appears to be a promising 3-D measurement solution to plant phenotyping for structural analysis.

## 7.2 Contributions

The main contributions and novel aspects of this thesis are summarised as follows:

- An overall system for plant phenotyping was proposed and its performances were fully analysed and evaluated.

- A low-cost multi-view 2-D image acquisition method was proposed to deal with the plant's complex shape and self-occlusion problem, which only used off-the-shelf equipments and resources (digital camera, turn-table and free software library) to capture high quality 2-D images. Based on the proposed method, tens of thousands of plant images were captured without pre-calibration, and a huge multi-view image dataset of diverse plants was built up, which covering *Arabidopsis thaliana*, oat, wheat, maize, foragegrass (*Lolium*), clover, Physalis, brassica, etc.

- A camera parameters and pose estimation method was developed, which includes the following steps: i) extraction of local invariant features for each image; ii) feature matching from neighboring overlapped image pairs, and iii) estimation of camera parameters and pose based on incremental Structure from Motion.

- An accurate multi-view stereo 3-D reconstruction method was proposed to generate a more dense, detailed and complete 3-D point cloud, which consists of stereo pair selection, depth maps computation, depth maps refinement and merging. The experimental results show that the proposed methods are highly reliable for phenotyping of various plants during their entire growth cycles.

- A novel approach to extract geometric information about plant organs was developed, which includes an automatic 3-D point cloud segmentation and a semi-automatic interactive estimation of branch angle directly from 3-D point cloud of plant. Experimental results show that the proposed approach is efficient and reliable, and appears to be a promising 3-D measurement solution to plant phenotyping for structural analysis.

- For camera parameter estimation, some local invariant features, e.g. SIFT, ASIFT, AKaze, DAISY, were used to make a comparison analysis. I found that: i) DAISY and AKAZE can produce many more 3-D matching keypoints than SUFR/SIFT, but 3-D keypoints produced by DAISY are mainly located in planar regions like leaves of plant or ruler (as reference), and 3-D keypoints detected by AKAZE focused on the boundary or edges of plant or ruler; ii) DAISY and AKAZE can better estimate the camera pose than SURF/SIFT in both the case of wide baseline where the rotation angle is bigger (in other words, the number of captured images is small) and the case of texture-less plants.

- For multi-view 2-D image acquisition, I found that rotation steps of 3∼6 degree increments are optimal and feature matching does not perform well with rotation steps of more than 9∼10 degrees.

## 7.3 List of Publications

A list of publications resulting from this thesis is shown as follows:

(1) Lu Lou, Yonghuai Liu, Minglan Sheng, Jiwan Han, Fiona Corke and John H. Doonan. Estimation of branch angle from 3D point cloud of plants. *in Proceedings of International Conference of 3D Vision* (3DV 2015), pp. 554-561. Lyon, France, 2015.

(2) Lu Lou, Yonghuai Liu, Jiwan Han and John H. Doonan. Accurate Multi-View Stereo 3D Reconstruction for Cost-effective Plant Phenotyping. *in Proceedings of International Conference on Image Analysis and Recognition* (ICIAR 2014), pp. 349-356. Algarve, Portugal, 2014.

(3) Lu Lou, Yonghuai Liu, Minglan Sheng, Jiwan Han and John H. Doonan. A cost-effective automatic 3D reconstruction pipeline for plants using multi-view images. *in Proceedings of the 15th Towards Autonomous Robotic Systems conference* (TAROS 2014), pp. 221-230. Birmingham, UK, 2014.

(4) Lu Lou, Suzana Barreto, Rokas Zmuidzinavicius, Mark James Neal, Hannah M. Dee, Frédéric Labrosse. Vision-Aided IMU Estimation of Attitude and Orientation for a Driverless Car. *in Proceedings of the 13th Towards Autonomous Robotic Systems conference* (TAROS 2012), pp. 465-466. Bristol, UK, 2012.

(5) Lu Lou, Mark James Neal, Frédéric Labrosse, Juan Cao. An Approach to Improving Attitude Estimation Based on Low-Cost MEMS-IMU for Mobile Robot Navigation. *in Proceedings of the 12th Towards Autonomous Robotic Systems conference* (TAROS 2011), pp. 378-379. Sheffield, UK, 2011.

In addition, some research reports were orally presented in the following workshops:

(1) Lu Lou. 3D Reconstruction of plant from uncalibrated multi-view images. *International Workshop on Image Analysis Methods for the Plant Sciences* (CPIB 2013), University of Nottingham, UK, 2-3 September 2013.

(2) Lu Lou. 3D modelling and analysis of plants. *RIVIC Graduate School 2013*, Bangor University, UK, 20-21 July 2013.

## 7.4 Future Directions

Possible future directions for further development are outlined as below:

(1) For multi-view image acquisition system, acquisition is currently manual, but a full automated ability can significantly reduce the time required by setup and acquisition, and is therefore more attractive for biologists.

(2) When triggering the camera to take images under "continuous shooting mode", the shooting mode can damage the camera shutter. Moreover, I found the Nikon camera cannot work well under "continuous shooting mode" if the number of captured images exceed about 100. This problem is probably caused by the Nikon camera's limited data buffer or data write

speed. Hence, the acquisition through cameras video recording such be evaluated, and 3-D reconstruction from video data remains as important further research.

(3) Currently, the final 3-D model of plants consists of only point clouds without meshes so that it cannot be directly processed in the common commercial CAD/3D software. Although surface reconstruction for plants is still challenging problem, significant effort will be required to implement this task.

(4) Compared with the performance of the real-time 3-D sensing technique, the proposed 3-D reconstruction method is still relatively slow, especially when processing hundreds of images. Therefore, it is necessary to improve the speed of the proposed 3-D reconstruction method in the future work.

(5) The results of segmentation is crucial for extracting useful phenotypic features from plant 3-D pint cloud. Fully automated 3-D segmentation method is still a rather challenging problem, because a meaningful segmentation on a plant can be quite different in other plant species or morphological mutants. Hence, an efficient fully automated 3-D segmentation method could be developed, perhaps by the combination of machine learning, prior knowledge or user intervention in the future.

In addition, the web based framework should be an important further researching task, illustrated in Fig. 7.1.

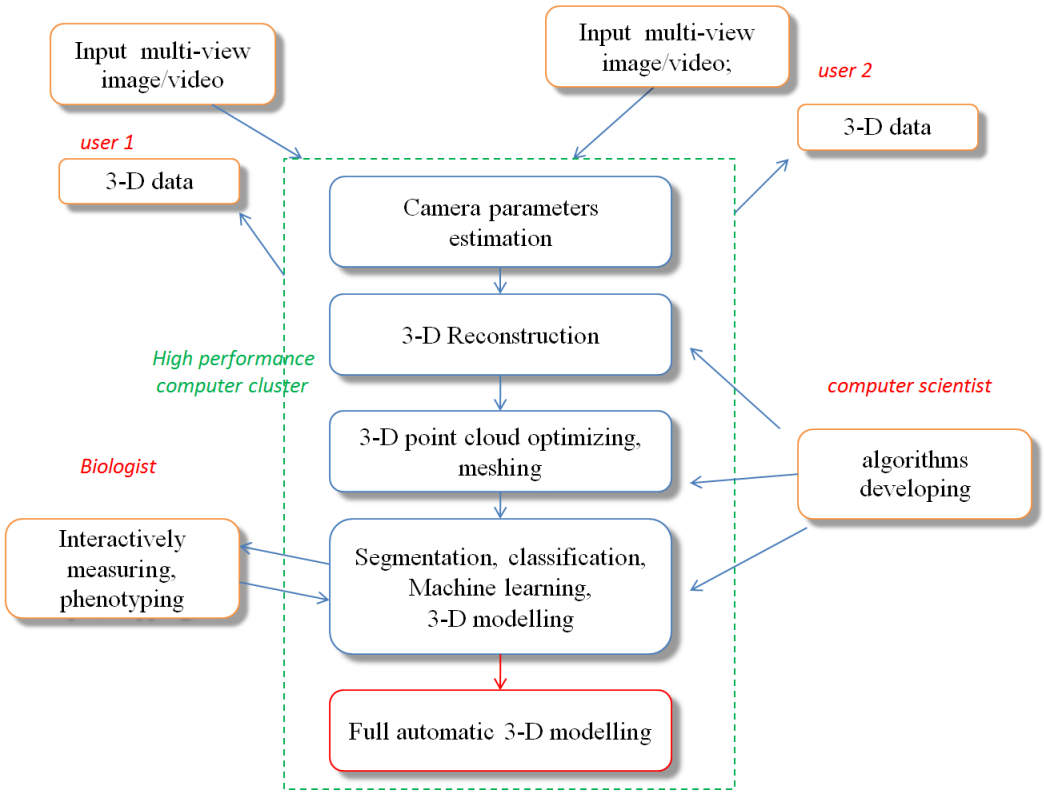



Figure 7.1: The web based framework of 3-D reconstruction for plants phenotyping and modelling in the future.

# References


3DSOM. 2014. *3D Modelling Software (3DSOM).* `http://www.3dsom.com/`.

Agarwal, Sameer, Snavely, Noah, Seitz, Steven M., & Szeliski, Richard. 2010. Bundle adjustment in the large. *Pages 29–42 of: the European Conference on Computer Vision.*

Agarwal, Sameer, Furukawa, Yasutaka, Snavely, Noah, Simon, Ian, Curless, Brian, Seitz, Steven M., & Szeliski, Richard. 2011. Building rome in a day. *Communications of the ACM*, **54**(10), 105–112.

Alcantarilla, Pablo, Nuevo, Jesus, & Bartoli, Adrien. 2013. Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces. *Pages 13.1–13.11 of: 2013 British Machine Vision Conference.*

Alenyà, Guillem, Dellen, Babette, & Torras, Carme. 2011. 3D modelling of leaves from color and ToF data for robotized plant measuring. *Pages 3408–3414 of: 2011 IEEE International Conference on Robotics and Automation (ICRA).*

Ann Arbor, Michigan. 2014. *Radial distortion of lens.* `http://www.uni-koeln.de/~al001/radcor_files/hs100.htm`.

Apelt, Federico, Breuer, David, Nikoloski, Zoran, Stitt, Mark, & Kragler, Friedrich. 2015. Phytotyping4D: a light-field imaging system for non-invasive and accurate monitoring of spatio-temporal plant growth. *The Plant Journal*, **82**(4), 693–706.

Artec. 2013. *Artec Eva 3D scanner.* `http://www.artec3d.com/hardware/artec-eva/`.

Arya, Sunil, Mount, David M., Netanyahu, Nathan S., Silverman, Ruth, & Wu, Angela Y. 1998. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *Journal of the ACM*, **45**(6), 891–923.

Bac, C. W., Hemming, J., & Henten, E. J. Van. 2013. Robust pixel-based classification of obstacles for robotic harvesting of sweet-pepper. *Computers and Electronics in Agriculture*, **96**(12), 148–162.

Backes, Andre Ricardo, Casanova, Dalcimar, & Bruno, Odemir Martinez. 2009. Plant leaf identification based on volumetric fractal dimension. *International Journal of Pattern Recognition and Artificial Intelligence*, **23**(6), 1145–1160.

Bailey, Regina. 2014. *Phenotype*. `http://biology.about.com/od/geneticsglossary/g/phenotype.htm`.

Barillot, Romain, Louarn, Gaëtan, Escobar-Gutiérrez, Abraham J., Huynh, Pierre, & Combes, Didier. 2011. How good is the turbid medium-based approach for accounting for light partitioning in contrasted grass–legume intercropping systems? *Annals of Botany*, **108**(6), 1013–1024.

Bay, Herbert, Ess, Andreas, Tuytelaars, Tinne, & Gool, Luc Van. 2008. Speeded-Up Robust Features (SURF). *Comput. Vis. Image Underst.*, **110**, 346–359.

Biskup, Bernhard, Scharr, Hanno, Schurr, Ulrich, & Rascher, Uwe. 2007. A stereo imaging system for measuring structural parameters of plant canopies. *Plant, Cell & Environment*, **30**(10), 1299–1308.

Boudon, Frédéric, Pradal, Christophe, Cokelaer, Thomas, Prusinkiewicz, Przemyslaw, & Godin, Christophe. 2012. L-Py: an L-system simulation framework for modeling plant architecture development based on a dynamic language. *Frontiers in Plant Science*, **3**(4), 279–286.

Bouget, Jean Yves. 2012. *Camera calibration toolbox for Matlab*. `http://www.vision.caltech.edu/bouguetj/calibdoc/`.

Bradley, Derek, Boubekeur, Tamy, & Heidrich, Wolfgang. 2010. Accurate multi-view reconstruction using robust binocular stereo and surface meshing. *Pages 1–8 of: IEEE Conference on Computer Vision & Pattern Recognition.*

Bruno, Odemir Martinez, de Oliveira Plotze, Rodrigo, Falvo, Mauricio, & de Castro, Mário. 2008. Fractal dimension applied to plant identification. *Information Sciences*, **178**(12), 2722–2733.

Campbell, Neill D., Vogiatzis, George, Hernández, Carlos, & Cipolla, Roberto. 2008. Using Multiple Hypotheses to Improve Depth-Maps for Multi-View Stereo. *Pages 766–779 of: the 10th European Conference on Computer Vision.*

Cao, Junjie, Tagliasacchi, Andrea, Olson, Matt, Zhang, Hao, & Su, Zhixun. 2010. Point cloud skeletons via laplacian based contraction. *Pages 187–197 of: 2010 Shape Modeling International Conference.* IEEE.

Carl, Nave. 2014. *Thin Lens Equation.* `http://hyperphysics.phy-astr.gsu.edu/hbase/geoopt/lenseq.html`.

Chen, Xiaobai, Golovinskiy, Aleksey, & Funkhouser, Thomas. 2009. A benchmark for 3D mesh segmentation. *ACM Transactions on Graphics (TOG)*, **28**(3), 341–352.

Cobb, J. N., DeClerck, G., Greenberg, A., Clark, R., & McCouch, S. 2013. Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. *Theoretical and Applied Genetics*, **126**, 867–887.

Database, DigitalCamera. 2015. *Digital Camera Database - One Stop Digicam Resource.* `http://www.digicamdb.com/`.

Dauzat, Jean, Clouvel, Pascal, Luquet, Delphine, & Martin, Pierre. 2008. Using virtual plants to analyse the light-foraging efficiency of a low-density cotton crop. *Annals of Botany*, **101**(8), 1153–1166.

Delaunoy, Amaël, Prados, Emmanuel, Piracés, Pau Gargallo I., Pons, Jean-Philippe, & Sturm, Peter. 2008. Minimizing the multi-view stereo reprojection error for triangular surface meshes. *Pages 1–10 of: 2008 British Machine Vision Conference.*

Dey, Debadeepta, Mummert, Lily, & Sukthankar, Rahul. 2012. Classification of plant structures from uncalibrated image sequences. *Pages 329–336 of: 2012 IEEE Workshop on Applications of Computer Vision (WACV).*

Diebel, James. 2006. Representing attitude: Euler angles, unit quaternions, and rotation vectors. *Matrix*, **58**, 15–16.

Dornbusch, Tino, Wernecke, Peter, & Diepenbrock, Wulf. 2007. A method to extract morphological traits of plant organs from 3D point clouds as a database for an architectural plant model. *Ecological modelling*, **200**(1), 119–129.

Dpreview. 2015. *Digitla Photography Review.* `http://www.dpreview.com/`.

Drouet, JL. 2003. MODICA and MODANCA: modelling the three-dimensional shoot structure of graminaceous crops from two methods of plant description. *Field Crops Research*, **83**(2), 215–222.

El-Omari, Samir, & Moselhi, Osama. 2011. Integrating automated data acquisition technologies for progress reporting of construction projects. *Automation in Construction*, **20**(6), 699–705.

Esteban, Carlos Hernández, & Schmitt, Francis. 2004. Silhouette and stereo fusion for 3D object modeling. *Computer Vision and Image Understanding*, **96**(3), 367–392.

Evers, Jochem B., Vos, Jan, Fournier, Christian, Andrieu, Bruno, Chelle, Michael, & Struik, Paul C. 2005. Towards a generic architectural model of tillering in Gramineae, as exemplified by spring wheat (Triticum aestivum). *New Phytologist*, **166**(3), 801–812.

Faugeras, Olivier. 1993. *Three-dimensional computer vision: a geometric viewpoint.* MIT press.

Faugeras, Olivier, & Keriven, Renaud. 1997. Level set methods and the stereo problem. *Pages 272–283 of: Scale-Space Theory in Computer Vision*. Springer.

Fernández, Roemi, Montes, Héctor, Salinas, Carlota, Sarria, Javier, & Armada, Manuel. 2013. Combination of rgb and multispectral imagery for discrimination of cabernet sauvignon grapevine elements. *Sensors*, **13**(6), 7838–7859.

Fischler, Martin A., & Bolles, Robert C. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, **24**(6), 381–395.

Flavel, R. J., Guppy, C. N., Tighe, M., Watt, M., Mcneill, A., & Young, I. M. 2012. Non-destructive quantification of cereal roots in soil using high-resolution X-ray tomography. *Journal of Experimental Botany*, **63**(7), 2503–2511.

France, INRA. 2014. *Phenopsis platform*. `http://www.plant-phenotyping-network.eu/eppn/inra_phenopsis`.

Furbank, Robert T., & Tester, Mark. 2011. Phenomics–technologies to relieve the phenotyping bottleneck. *Trends in Plant Science*, **16**(12), 635–644.

Furukawa, Yasutaka, & Ponce, Jean. 2010. Accurate, Dense, and Robust Multiview Stereopsis. *IEEE Trans. Pattern Anal. Mach. Intell.*, **32**, 1362–1376.

Gaëtan, Louarn, Serge, Carré, Annie, Eprinchard, Didier, Combes, & Frederic, Boudon. 2012. Characterization of whole plant leaf area properties using laser scanner point clouds. *Pages 250–253 of: 2012 IEEE Fourth International Symposium on Plant Growth Modeling, Simulation, Visualization and Applications*.

Garbout, Amin, Munkholm, Lars J., Hansen, Søren B., Petersen, Bjørn M., Munk, Ole L., & Pajor, Radoslaw. 2012. The use of PET/CT scanning technique for 3D visualization and quantification of real-time soil/plant interactions. *Plant and Soil*, **352**(1), 113–127.

GmbH, Raytrix. 2014. *Raytrix 3D light field cameras.* `http://www.raytrix.de/`.

Goesele, Michael, Curless, Brian, & Seitz, Steven M. 2006. Multi-View Stereo Revisited. *Pages 2402–2409 of: the 2006 IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2.

Goesele, Michael, Snavely, Noah, Curless, Brian, Hoppe, Hugues, & Seitz, Steven M. 2007. Multi-view stereo for community photo collections. *Pages 1–8 of: IEEE 11th International Conference on Computer Vision.*

Golovinskiy, Aleksey, & Funkhouser, Thomas. 2009. Consistent segmentation of 3D models. *Computers & Graphics*, **33**(3), 262–269.

Granier, C., Aguirrezabal, L., Chenu, K., Cookson, S. J., Dauzat, M., Hamard, P., Thioux, J. J., Rolland, G., Bouchier-Combaud, S., Lebaudy, A., Muller, B., Simonneau, T., & Tardieu, F. 2006. PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in Arabidopsis thaliana permitted the identification of an accession with low sensitivity to soil water deficit. *New Phytologist*, **169**(3), 623–635.

Granier, Christine, & Tardieu, Francois. 2009. Multi-scale phenotyping of leaf expansion in response to environmental changes: the whole is more than the sum of parts. *Plant, Cell & Environment*, **32**(9), 1175–1184.

Gregory, P. J., Hutchison, D. J., Read, D. B., Jenneson, P. M., Gilboy, W. B., & Morton, E. J. 2003. Non-invasive imaging of roots with high resolution X-ray micro-tomography. *Plant & Soil*, **255**(1), 351–359.

Guillot, Gilles, Lorén, Niklas, & Rudemo, Mats. 2009. Spatial prediction of weed intensities from exact count data and image-based estimates. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **58**(4), 525–542.

Habbecke, Martin, & Kobbelt, Leif. 2007. A surface-growing approach to multi-view stereo reconstruction. *Pages 1–8 of: IEEE Conference on Computer Vision and Pattern Recognition.*

Han, Jungong, Shao, Ling, Xu, Dong, & Shotton, Jamie. 2013. Enhanced computer vision with microsoft kinect sensor: A review. *IEEE Transactions on Cybernetics*, **43**(5), 1318–1334.

Hartley, R. I., & Zisserman, A. 2004. *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Hartmann, Anja, Czauderna, Tobias, Hoffmann, Roberto, Stein, Nils, & Schreiber, Falk. 2011. HTPheno: an image analysis pipeline for high-throughput plant phenotyping. *BMC Bioinformatics*, **12**(1), 148.

He, D. X., Matsuura, Y., Kozai, T., & Ting, K. C. 2003. Binocular Stereovision System for Transplant Growth Variables Analysis. *Applied Engineering in Agriculture*, **19**(5), 611–618.

Hirschmuller, H. 2008. Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **30**(2), 328–341.

Hopkinson, Chris, Chasmer, Laura, Young-Pow, Colin, & Treitz, Paul. 2004. Assessing forest metrics with a ground-based scanning lidar. *Canadian Journal of Forest Research*, **34**(3), 573–583.

Hosoi, F., Yoshimi, K., Shimizu, Y., & Omasa, K. 2005. 3-D measurement of trees using a portable scanning lidar. *Phyton*, **45**(4), 497–500.

Hosoi, Fumiki, & Omasa, Kenji. 2006. Voxel-based 3-D modeling of individual trees for estimating leaf area density using high-resolution portable scanning lidar. *IEEE Transactions on Geoscience and Remote Sensing*, **44**(12), 3610–3618.

Hosoi, Fumiki, & Omasa, Kenji. 2009. Estimating vertical plant area density profile and growth parameters of a wheat canopy at different growth stages using three-dimensional portable lidar imaging. *ISPRS Journal of Photogrammetry and Remote Sensing*, **64**(2), 151–158.

Houle, David, Govindaraju, Diddahally R., & Omholt, Stig. 2010. Phenomics: the next challenge. *Nature Reviews Genetics*, **11**(12), 855–866.

Humplík, Jan F, Lazár, Dušan, Husičková, Alexandra, & Spíchal, Lukáš. 2015. Automated phenotyping of plant shoots using imaging methods for analysis of plant stress responses - a review. *Plant methods*, **11**(1), 1–10.

ImageJ. 2014. *Image processing in Java*. `http://rsbweb.nih.gov/ij/`.

Initiative, Arabidopsis Genome. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*, **408**(6814), 796–815.

ISAAA, International Service For the Acquisition of AGRI-Biotech Applications. 2006. *Conventional Plant Breeding*. `https://isaaa.org/resources/publications/pocketk/13/default.asp`.

Ivanov, N., Boissard, P., Chapron, M., & Valery, P. 1994. Estimation of the height and angles of orientation of the upper leaves in the maize canopy using stereovision. *Agronomie*, **14**(3), 183–194.

Ivanov, N., Boissard, P., Chapron, M., & Andrieu, B. 1995. Computer stereo plotting for 3-D reconstruction of a maize canopy. *Agricultural and Forest Meteorology*, **75**(1), 85–102.

Jancosek, M., & Pajdla, T. 2009. Segmentation based multi-view stereo. *In: Computer Vision Winter Workshop 2009*. PRIP, Vienna University of Technology, Austria.

Jancosek, M., & Pajdla, T. 2011. Multi-view Reconstruction Preserving Weakly-supported Surfaces. *Pages 3121–3128 of: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Johannsen, Wilhelm. 1903. Erblichkeit in Populationen und in reinen Linien. *Jena: Gustav Fischer Verlag*.

Johannsen, Wilhelm. 1911. The genotype conception of heredity. *Am Nat*, **45**, 129–159.

Johnson, Marc T. J., & Agrawal, Anurag A. 2005. Plant genotype and environment interact to shape a diverse arthropod community on evening primrose (Oenothera biennis). *Ecology*, **86**(4), 874–885.

Jongerius, Jerry. 2012. *How to properly use the "Thin Lens Formula"*. `http://www.panohelp.com/thinlensformula.html`.

Kalogerakis, Evangelos, Hertzmann, Aaron, & Singh, Karan. 2010. Learning 3D mesh segmentation and labeling. *Acm Transactions on Graphics*, **29**(4), 157–166.

Kaminuma, Eli, Heida, Naohiko, Tsumoto, Yuko, Yamamoto, Naoki, Goto, Nobuharu, Okamoto, Naoki, Konagaya, Akihiko, Matsui, Minami, & Toyoda, Tetsuro. 2004. Automatic quantification of morphological traits via three-dimensional measurement of Arabidopsis. *The Plant Journal*, **38**(2), 358–365.

Karunakaran, C., Jayas, D. S., & White, N. D. G. 2004. Detection of internal wheat seed infestation by Rhyzopertha dominica using X-ray imaging. *Journal of Stored Products Research*, **40**(5), 507–516.

Karwowski, Radoslaw, & Prusinkiewicz, Przemyslaw. 2003. Design and implementation of the L+C modeling language. *Electronic Notes in Theoretical Computer Science*, **86**(2), 134–152.

Khoshelham, Kourosh, & Elberink, Sander Oude. 2012. Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors*, **12**(2), 1437–1454.

Kinect. 2011. *Kinect for Windows*. `https://www.microsoft.com/en-us/kinectforwindows/meetkinect/features.aspx`.

Klose, Ralph, Penlington, Jaime, & Ruckelshausen, Arno. 2009. Usability study of 3D time-of-flight cameras for automatic plant phenotyping. *Applied Sciences*, **69**(12), 93–105.

Kniemeyer, Ole, & Kurth, Winfried. 2007. The modelling platform GroIMP and the programming language XL. *Pages 570–572 of: Third International Symposium on Applications of Graph Transformations with Industrial Relevance*.

Kolev, Kalin, & Cremers, Daniel. 2008. Integration of multiview stereo and silhouettes via convex functionals on convex domains. *Pages 752–765 of: European Conference on Computer Vision*.

Kolev, Kalin, Klodt, Maria, Brox, Thomas, & Cremers, Daniel. 2009. Continuous global optimization in multiview 3d reconstruction. *International Journal of Computer Vision*, **84**(1), 80–96.

Kolmogorov, Vladimir, & Zabih, Ramin. 2001. Multi-camera Scene Reconstruction via Graph Cuts. *Pages 82–96 of: European Conference on Computer Vision*.

Konishi, Atsumi, Eguchi, Akira, Hosoi, Fumiki, & Omasa, Kenji. 2009. 3D monitoring spatio–temporal effects of herbicide on a whole plant using combined range and chlorophyll a fluorescence imaging. *Functional Plant Biology*, **36**(11), 874–879.

Krämer, U. 2015. Planting molecular functions in an ecological context with Arabidopsis thaliana. *Elife Sciences*, **4**(4), 1–13.

Lanman, Douglas, & Taubin, Gabriel. 2009. *Build Your Own 3D Scanner: Optical Triangulation for Beginners*. `http://rsbweb.nih.gov/ij/`.

LemnaTec. 2014. *Image process in Biology*. `http://www.lemnatec.com`.

Levoy, Marc. 2002. *Stanford Spherical Gantry*. `https://graphics.stanford.edu/projects/gantry/`.

Lhuillier, Maxime, & Quan, Long. 2005. A quasi-dense approach to surface reconstruction from uncalibrated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**(3), 418–433.

Li, Dawei, Xu, Lihong, Tan, Chengxiang, Goodman, Erik D, Fu, Daichang, & Xin, Longjiao. 2015. Digitization and Visualization of Greenhouse Tomato Plants in Indoor Environments. *Sensors*, **15**(2), 4019–4051.

Li, Han, Lee, Won Suk, & Wang, Ku. 2014a. Identifying blueberry fruit of different growth stages using natural outdoor color images. *Computers and Electronics in Agriculture*, **106**, 91–101.

Li, Ji. 2014. *3D machine vision system for robotic weeding and plant phenotyping.* Ph.D. thesis.

Li, L., Zhang, Q., & Huang, D. 2014b. A review of imaging techniques for plant phenotyping. *Sensors*, **14**(11), 20078–20111.

Li, Yangyan, Fan, Xiaochen, Mitra, Niloy J., Chamovitz, Daniel, Cohen-Or, Daniel, & Chen, Baoquan. 2013. Analyzing growing plants from 4D point cloud data. *Acm Transactions on Graphics*, **32**(6), 1–10.

Lindenmayer, Aristid. 1968. Mathematical models for cellular interactions in development I. Filaments with one-sided inputs. *Journal of theoretical biology*, **18**(3), 280–299.

Liu, Rong, & Zhang, Hao. 2004. Segmentation of 3D meshes through spectral clustering. *Pages 298–305 of: Proceedings on 12th Pacific Conference on Computer Graphics and Applications.*

Liu, Yebin, Cao, Xun, Dai, Qionghai, & Xu, Wenli. 2009. Continuous depth estimation for multi-view stereo. *Pages 2121–2128 of: IEEE Conference on Computer Vision & Pattern Recognition.*

Lourakis, Manolis I. A., & Argyros, Antonis A. 2009. SBA: A software package for generic sparse bundle adjustment. *ACM Transactions on Mathematical Software (TOMS)*, **36**(1), 1–30.

Lowe, David G. 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, **60**, 91–110.

Merrell, Paul, Akbarzadeh, Amir, Wang, Liang, michael Frahm, Jan, & Nister, Ruigang Yang David. 2007. Real-time visibility-based fusion of depth maps. *Pages 1–8 of: IEEE 11th International Conference on Computer Vision (ICCV 2007).*

MeshLab. 2014. *MeshLab Software v1.3.3.* `http://meshlab.sourceforge.net/`.

Meyer, George E., Neto, Joao Camargo, Jones, David D., & Hindman, Timothy W. 2004. Intensified fuzzy clusters for classifying plant, soil, and

residue regions of interest from color images. *Computers and electronics in agriculture*, **42**(3), 161–180.

Montalvo, M., Guerrero, José Miguel, Romeo, J., Emmi, Luis, Guijarro, María, & Pajares, Gonzalo. 2013. Automatic expert system for weeds/crops identification in images from maize fields. *Expert Systems with Applications*, **40**(1), 75–82.

Moreels, Pierre, & Perona, Pietro. 2007. Evaluation of features detectors and descriptors based on 3d objects. *International Journal of Computer Vision*, **73**(3), 263–284.

Morel, Jean-Michel, & Yu, Guoshen. 2009. ASIFT: A New Framework for Fully Affine Invariant Image Comparison. *SIAM J. Img. Sci.*, **2**(2), 438–469.

Moulon, Pierre. 2012. *openMVG: open Multiple View Geometry.* `http://imagine.enpc.fr/~moulonp/openMVG/`.

Mündermann, Lars, Erasmus, Yvette, Lane, Brendan, Coen, Enrico, & Prusinkiewicz, Przemyslaw. 2005. Quantitative modeling of Arabidopsis development. *Plant physiology*, **139**(2), 960–968.

Nandi, Chandra Sekhar, Tudu, Bipan, & Koley, Chiranjib. 2014. A Machine Vision-Based Maturity Prediction System for Sorting of Harvested Mangoes. *IEEE Transactions on Instrumentation and Measurement*, **63**(7), 1722–1730.

Nistér, David. 2004. An Efficient Solution to the Five-Point Relative Pose Problem. *IEEE Trans. Pattern Anal. Mach. Intell.*, **26**(6), 756–777.

Omasa, Kenji, Hosoi, Fumiki, & Konishi, Atsumi. 2007. 3D lidar imaging for detecting and understanding plant responses and canopy structure. *Journal of experimental botany*, **58**(4), 881–898.

Onyango, Christine M., & Marchant, J. A. 2003. Segmentation of row crop plants from weeds using colour and morphology. *Computers and electronics in agriculture*, **39**(3), 141–155.

Onyango, Christine M., & Marchant, John A. 2001. Physics-based colour image segmentation for scenes containing vegetation and soil. *Image and vision computing*, **19**(8), 523–538.

O'Rourke, Joseph. 1985. Finding minimal enclosing boxes. *International journal of computer & information sciences*, **14**(3), 183–199.

Paproki, Anthony, Sirault, Xavier, Berry, Scott, Furbank, Robert, & Fripp, Jurgen. 2012. A novel mesh processing based technique for 3D plant analysis. *BMC plant biology*, **12**(1), 63.

Paulus, Stefan, Dupuis, Jan, Mahlein, Anne-Katrin, & Kuhlmann, Heiner. 2013. Surface feature based classification of plant organs from 3D laser-scanned point clouds for plant phenotyping. *BMC bioinformatics*, **14**(1), 238.

Paulus, Stefan, Dupuis, Jan, Riedel, Sebastian, & Kuhlmann, Heiner. 2014a. Automated analysis of barley organs using 3D laser scanning: An approach for high throughput phenotyping. *Sensors*, **14**(7), 12670–12686.

Paulus, Stefan, Schumann, Henrik, Kuhlmann, Heiner, & Léon, Jens. 2014b. High-precision laser scanning system for capturing 3D plant architecture and analysing growth of cereal plants. *Biosystems Engineering*, **121**, 1–11.

Paulus, Stefan, Behmann, Jan, Mahlein, Anne-Katrin, Plümer, Lutz, & Kuhlmann, Heiner. 2014c. Low-cost 3D systems: suitable tools for plant phenotyping. *Sensors*, **14**(2), 3001–3018.

Photometadata. 2015. *Exchangeable Image File Format*. `http://www.photometadata.org/meta-resources-metadata-types-standards-exif`.

PlantEye. 2015. *PlantEye: a 3D laser scanner for plants*. `http://phenospex.com/`.

Pollefeys, Marc. 2002. *Radial distortion*. `http://www.cs.unc.edu/~marc/tutorial/node41.html`.

Pons, Jean-Philippe, Keriven, Renaud, & Faugeras, Olivier. 2005. Modelling dynamic scenes by registering multi-view image sequences. *Pages 822–827 of: IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2. IEEE.

Pound, Michael P., French, Andrew P., Murchie, Erik H., & Pridmore, Tony P. 2014. Automated recovery of three-dimensional models of plant shoots from multiple color images. *Plant physiology*, **166**(4), 1688–1698.

Pradal, Christophe, Dufour-Kowalski, Samuel, Boudon, Frédéric, Fournier, Christian, & Godin, Christophe. 2008. OpenAlea: a visual programming and component-based software platform for plant modelling. *Functional plant biology*, **35**(10), 751–760.

Pradal, Christophe, Boudon, Frédéric, Nouguier, Christophe, Chopard, Jérôme, & Godin, Christophe. 2009. PlantGL: a Python-based geometric library for 3D plant modelling at different scales. *Graphical models*, **71**(1), 1–21.

Prusinkiewicz, P. W., Remphrey, W. R., Davidson, C. G., & Hammel, M. S. 1994. Modeling the architecture of expanding Fraxinus pennsylvanica shoots using L-systems. *Canadian Journal of Botany*, **72**(5), 701–714.

Prusinkiewicz, Przemyslaw, & Hanan, Jim. 1990. Visualization of botanical structures and processes using parametric L-systems. *European Journal of Cancer*, **50**(15), 2570–2582.

Prusinkiewicz, Przemyslaw, Hanan, Jim, & Měch, Radomír. 1999. An L-system-based plant modeling language. *Pages 395–410 of: International Workshop on Applications of Graph Transformations with Industrial Relevance (AGTIVE1999)*.

Quan, L., Tan, P., Zeng, G., Yuan, L., Wang, J. D., & Kang, S. B. 2006. Image-based plant modeling. *Acm Transactions on Graphics*, **25**, 599–604.

ReconstructMe. 2014. *ReconstructMe V2.4*. `http://reconstructme.net/`.

Reuzeau, C., Frankard, V., Hatzfeld, Y., Sanz, A., Camp, W. V., Lejeune, P., Wilde, C. D., Lievens, K., Wolf, J. D., & Vranken, E. 2006. TraitMill

(TM): A high throughput functional genomics platform for the phenotypic analysis of cereals. *Plant Genetic Resources*, **4**(1), 20–24.

Romeo, J., Pajares, Gonzalo, Montalvo, M., Guerrero, José Miguel, Guijarro, María, & De La Cruz, J. M. 2013. A new Expert System for greenness identification in agricultural images. *Expert Systems with Applications*, **40**(6), 2275–2286.

Rusu, Radu Bogdan, Blodow, Nico, Marton, Zoltan Csaba, & Beetz, Michael. 2008. Aligning point cloud views using persistent feature histograms. *Pages 3384–3391 of: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2008)*.

Samal, Ashok, Brandle, James R., & Zhang, Dongsheng. 2006. Texture as the basis for individual tree identification. *Information Sciences*, **176**(5), 565–576.

Santos, Thiago T., & Oliveira, Alberto A. D. 2012. Image-based 3D digitizing for plant architecture analysis and phenotyping. *Pages 21–28 of: The Workshop on Industry Applications in SIBGRAPI*, vol. 2012.

Schauer, Nicolas, & Fernie, Alisdair R. 2006. Plant metabolomics: towards biological function and mechanism. *Trends in plant science*, **11**(10), 508–516.

Schilling, Christophe H., Edwards, Jeremy S., & Palsson, Bernhard O. 1999. Toward metabolic phenomics: analysis of genomic data using flux balances. *Biotechnology Progress*, **15**(3), 288–295.

Schulze, Waltraud X., & Usadel, Björn. 2010. Quantitation in mass-spectrometry-based proteomics. *Annual review of plant biology*, **61**, 491–516.

Schurr, U., Heckenberger, U., Herdel, K., Walter, A., & Feil, R. 2000. Leaf development in Ricinus communis during drought stress: dynamics of growth processes, of cellular structure and of sink–source transition. *Journal of Experimental Botany*, **51**(350), 1515–1529.

Seitz, Steven M., Curless, Brian, Diebel, James, Scharstein, Daniel, & Szeliski, Richard. 2006a. A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms. *Pages 519–528 of: 2006 IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1.

Seitz, Steven M., Curless, Brian, Diebel, James, Scharstein, Daniel, & Szeliski, Richard. 2006b. *Middlebury multi-view stereo dataset*. `http://vision.middlebury.edu/mview/data/`.

Shamir, Ariel. 2006. Segmentation and shape extraction of 3D boundary meshes. *State-of-the-Art Report, Proceedings Eurographics*, **2006**, 137–49.

Shan, Qi, Wu, Changchang, Curless, Brian, Furukawa, Yasutaka, Hernandez, Carlos, & Seitz, Steven M. 2014. Accurate geo-registration by ground-to-aerial image matching. *Pages 525–532 of: 2nd International Conference on 3D Vision (3DV 2014)*, vol. 1. IEEE.

Shlafman, Shymon, Tal, Ayellet, & Katz, Sagi. 2002. Metamorphosis of polyhedral surfaces using decomposition. *Pages 219–228 of: Computer Graphics Forum*, vol. 21. Wiley Online Library.

Sievänen, Risto, Nikinmaa, Eero, Nygren, Pekka, Ozier-Lafontaine, Harry, Perttunen, Jari, & Hakula, Harri. 2000. Components of functional-structural tree models. *Annals of forest science*, **57**(5), 399–412.

Sinha, Sudipta N., Mordohai, Philippos, & Pollefeys, Marc. 2007. Multi-view stereo via graph cuts on the dual of an adaptive tetrahedral mesh. *Pages 1–8 of: IEEE 11th International Conference on Computer Vision (ICCV 2007)*.

Sinoquet, Hervé, & Rivet, Pierre. 1997. Measurement and visualization of the architecture of an adult tree based on a three-dimensional digitising device. *Trees*, **11**(5), 265–270.

Sinoquet, Hervé, Rivet, Pierre, & Godin, Christophe. 1997. Assessment of the three-dimensional architecture of walnut trees using digitising. *Silva Fennica*, **31**(3).

Sleper, David Allen, & Poehlman, John Milton. 2006. Breeding field crops: fifth edition. *Journal of Sat Agricultural Research*, **2**(1).

Slicer, 3D. 2014. *3D Slicer V4.4*. `http://www.slicer.org/`.

Snavely, N., Simon, I., Goesele, M., & Szeliski, R. 2010. Scene Reconstruction and Visualization From Community Photo Collections. *Proceedings of the IEEE*, **98**(8), 1370–1390.

Snavely, Noah, Seitz, Steven M., & Szeliski, Richard. 2006. Photo tourism: exploring photo collections in 3D. *ACM Transactions on Graphics (TOG)*, **25**(3), 835–846.

Snavely, Noah, Seitz, Steven M., & Szeliski, Richard. 2008. Modeling the World from Internet Photo Collections. *Int. J. Comput. Vision*, **80**(2), 189–210.

Stephens, Chris Harris, & Mike. 1988. A combined corner and edge detector. *Pages 5–10 of: Fourth Alvey Vision Conference*.

Sticklen, Mariam B. 2007. Feedstock crop genetic engineering for alcohol fuels. *Crop science*, **47**(6), 2238–2248.

Strecha, Christoph, Fransens, Rik, & Van Gool, Luc. 2006. Combined Depth and Outlier Estimation in Multi-View Stereo. *Pages 2394–2401 of: 2006 IEEE Conference on Computer Vision and Pattern Recognition*.

Swain, Kishore C., Nørremark, Michael, Jørgensen, Rasmus N., Midtiby, Henrik S., & Green, Ole. 2011. Weed identification using an automated active shape matching (AASM) technique. *biosystems engineering*, **110**(4), 450–457.

Systems, Breeze. 2014. *DSLR Remote Pro Multi-Camera v3.3*. `http://www.breezesys.com/DSLRRemotePro/`.

Szeliski, Richard. 2010. *Computer vision: algorithms and applications*. Springer Science & Business Media.

Tardieu, F., and Schurr U. 2009. *White paper of Plant Phenotyping*. `http://www.plantphenomics.com/phenotyping2009`.

Tellaeche, Alberto, Burgos-Artizzu, Xavier P., Pajares, Gonzalo, & Ribeiro, Angela. 2008. A vision-based method for weeds identification through the Bayesian decision theory. *Pattern Recognition*, **41**(2), 521–530.

Tester, Mark, & Langridge, Peter. 2010. Breeding technologies to increase crop production in a changing world. *Science*, **327**(5967), 818–822.

Tola, Engin, Lepetit, Vincent, & Fua, Pascal. 2010. DAISY: An Efficient Dense Descriptor Applied to Wide-Baseline Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **32**(5), 815–830.

Tran, Son, & Davis, Larry. 2006. 3D surface reconstruction using graph cuts with surface constraints. *Pages 219–231 of: European Conference on Computer Vision*.

Triggs, Bill, McLauchlan, Philip F., Hartley, Richard I., & Fitzgibbon, Andrew W. 1999. Bundle adjustment - a modern synthesis. *Pages 298–372 of: International Workshop on Vision Algorithms: Theory and Practice*.

Ullman, Shimon. 1979. The interpretation of structure from motion. *Proceedings of the Royal Society of London B: Biological Sciences*, **203**(1153), 405–426.

Veeken, M. M. J. J. van der. 2006. *Automated corn plant spacing measurement at early growth stages using active computer vision*. Ph.D. thesis.

Vogiatzis, George, Hernández, Carlos, Torr, Philip H. S., & Cipolla, Roberto. 2007. Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **29**(12), 2241–2246.

Von Luxburg, Ulrike. 2007. A tutorial on spectral clustering. *Statistics and computing*, **17**(4), 395–416.

Vos, J., Marcelis, L. F. M., & Evers, J. B. 2007. Functional-structural plant modelling in crop production: adding a dimension. *Pages 1–12 of: the Frontis Workshop on Functional-Structural Plant Modelling in Crop Production*, vol. 22.

Vos, J., Evers, J. B., Bucksorlin, G. H., Andrieu, B, Chelle, M., & de Visser, P. H. 2010. Functional-structural plant modelling: a new versatile tool in crop science. *Journal of experimental botany*, **61**(8), 2101–2115.

Walter, Achim, Liebisch, Frank, & Hund, Andreas. 2015. Plant phenotyping: from bean weighing to image analysis. *Plant Methods*, **11**(1), 1–11.

Weiss, Ulrich, Biber, Peter, Laible, Stefan, Bohlmann, Karsten, & Zell, Andreas. 2010. Plant species classification using a 3D lidar sensor and machine learning. *Pages 339–345 of: 2010 Ninth International Conference on Machine Learning and Applications (ICMLA)*.

WinDIAS. 2014. *WinDIAS leaf image analysis system.* `http://www.delta-t.co.uk/product-display.asp?id=WD3%20Product&div=Plant%20Science`.

Winkelbach, Simon, Molkenstruck, Sven, & Wahl, Friedrich M. 2006. Low-cost laser range scanner and fast surface registration approach. *Pages 718–728 of: Pattern Recognition*, vol. 4174. Springer.

WPS. 2014. *Next Generation Digital Phenotyping.* `http://www.wps.eu/en`.

Wu, Changchang. 2011. *SiftGPU: A GPU implementation of scale invariant feature transform (SIFT).* `http://cs.unc.edu/~ccwu/siftgpu`.

Wu, Changchang. 2013. Towards linear-time incremental structure from motion. *Pages 127–134 of: International Conference on 3D Vision (3DV 2013)*.

Wu, Changchang. 2014. Critical configurations for radial distortion self-calibration. *Pages 25–32 of: IEEE Conference on Computer Vision and Pattern Recognition*.

Wu, Changchang, Agarwal, S., Curless, B., & Seitz, S. M. 2011. Multicore bundle adjustment. *Pages 3057–3064 of: 2011 IEEE Conference on Computer Vision and Pattern Recognition*.

Xu, Guili, Zhang, Fengling, Shah, Syed Ghafoor, Ye, Yongqiang, & Mao, Hanping. 2011. Use of leaf color images to identify nitrogen and potassium deficient tomatoes. *Pattern Recognition Letters*, **32**(11), 1584–1590.

Yamazaki, Shuntaro, Narasimhan, Srinivasa G., Baker, Simon, & Kanade, Takeo. 2009. The theory and practice of coplanar shadowgram imaging for acquiring visual hulls of intricate objects. *International Journal of Computer Vision*, **81**(3), 259–280.

Yang, Wanneng, Xu, Xiaochun, Duan, Lingfeng, Luo, Qingming, Chen, Shangbin, Zeng, Shaoqun, & Liu, Qian. 2011. High-throughput measurement of rice tillers using a conveyor equipped with x-ray computed tomography. *Review of Scientific Instruments*, **82**(2), 025102–025109.

Zach, Christopher, Pock, Thomas, & Bischof, Horst. 2007. A globally optimal algorithm for robust TV-L 1 range image integration. *Pages 1–8 of: IEEE 11th International Conference on Computer Vision.*

Zaharescu, Andrei, Boyer, Edmond, & Horaud, Radu. 2007. TransforMesh: A Topology-Adaptive Mesh-Based Approach to Surface Evolution. *Pages 166–175 of: 2007 Asian Conference on Computer Vision.*

Zande, Dimitry Van Der, Hoet, Wouter, Jonckheere, Inge, Aardt, Jan Van, & Coppin, Pol. 2007. Influence of measurement set-up of ground-based LiDAR for derivation of tree structure. *Agricultural and Forest Meteorology*, **141**(2), 147–160.

Zeng, Qingbing, Miao, Yubin, Liu, Chengliang, & Wang, Shiping. 2009. Algorithm based on marker-controlled watershed transform for overlapping plant fruit segmentation. *Optical Engineering*, **48**(2), 027201–027211.

Zhang, Zhengyou. 2000. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(11), 1330–1334.

Zheng, Bangyou, Shi, Lijuan, Ma, Yuntao, Deng, Qiyun, Li, Baoguo, & Guo, Yan. 2008. Comparison of architecture among different cultivars of hybrid

rice using a spatial light model based on 3-D digitising. *Functional Plant Biology*, **35**(10), 900–910.

Zheng, Liying, Zhang, Jingtao, & Wang, Qianyu. 2009. Mean-shift-based color segmentation of images containing green vegetation. *Computers and Electronics in Agriculture*, **65**(1), 93–98.